

7 March 2008 | S10

# Science



POLLARD STREET F



## COVER

An example of “art” by self-styled guerrilla artist Banksy, as seen in East London in November 2007. Human behavior that would be characterized as antisocial punishment can also be called art; prosocial institutions, most notably the campaign Keep Britain Tidy, refer to Banksy’s work as vandalism. See [page 1362](#).

*Photo: Chris Jackson/Getty Images*

## DEPARTMENTS

- 1301 [Science Online](#)
- 1303 [This Week in Science](#)
- 1309 [Editors’ Choice](#)
- 1312 [Contact Science](#)
- 1315 [Random Samples](#)
- 1317 [Newsmakers](#)
- 1414 [New Products](#)
- 1415 [Science Careers](#)

## EDITORIAL

- 1307 [A New Editor-in-Chief](#)  
*by Bruce Alberts*

## NEWS OF THE WEEK

- Ecologists Report Huge Storm Loses in China’s Forests 1318
- Brazilian Scientists Battle Animal Experimentation Bans 1319
- Test of Hawking’s Prediction on the Horizon With Mock ‘White Hole’ 1321

>> [Report p. 1367](#)

### SCIENCESCOPE 1321

- NSF Delays Three Projects to Get Better Handle on Costs 1322
- U.S. Biomedicine’s Mother Ship Braces for Lab Closings 1324
- Antimatter Experiment May Be Too Costly for NASA to Launch 1324
- Electron Shadow Hints at Invisible Rings Around a Moon 1325

>> [Report p. 1380](#)

## NEWS FOCUS

- Preparing for Doomsday: The State of Our Planet’s Defenses 1326
- Experts Find No Evidence for a Mammoth-Killer Impact 1331
- Corn Genomics Pops Wide Open 1333



1326

## LETTERS

- Editorial Expression of Concern 1335  
*D. Kennedy and B. Alberts*
- Rewarding Reviewers *M. A. Metz*
- Rating Reviewers *G. Marchionini*
- Preventing Inequity in International Research  
*I. Rudan* Response *C. A. Nelson III et al.*
- AIDS Vaccine Research: Consider Co-Infections  
*R. Gruters and A. Osterhaus*

## CORRECTIONS AND CLARIFICATIONS 1336

## BOOKS ET AL.

- The Second Life Herald** The Virtual Tabloid That Witnessed the Dawn of the Metaverse  
*P. Ludlow and M. Wallace, reviewed by D. Hunter* 1338
- HIV/AIDS, Illness, and African Well-Being**  
*T. Falola and M. M. Heaton, Eds., reviewed by A. G. Abimiku* 1339

## POLICY FORUM

- Moving Toward Transparency of Clinical Trials  
*D. A. Zarin and T. Tse* 1340

## PERSPECTIVES

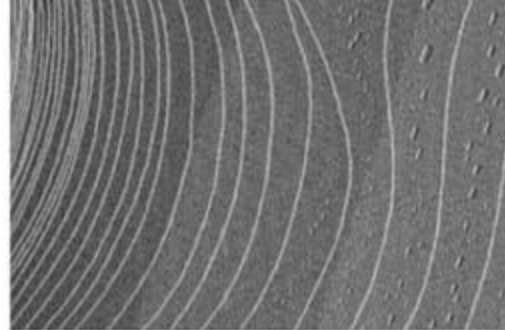
- Canyon Cutting on a Grand Time Scale  
*T. Atkinson and M. Leeder* >> [Report p. 1377](#) 1343
- A Tall Tale for U  
*M. Wickens and J. E. Kwak* 1344
- Punishment and Cooperation  
*H. Gintis* >> [Research Article p. 1362](#) 1345
- On Phytoplankton Trends  
*V. Smetacek and J. E. Cloern* 1346
- Sweet Conundrum  
*M. J. Birnbaum* >> [Report p. 1402](#) 1348
- Science 2.0  
*B. Shneiderman* 1349
- Retrospective: [Joshua Lederberg \(1925–2008\)](#)  
*S. S. Morse* 1351

## GORDON RESEARCH CONFERENCES 1310 & 1409



1338

[CONTENTS continued >>](#)



## SCIENCE EXPRESS

[www.sciencexpress.org](http://www.sciencexpress.org)

### CHEMISTRY

**Linked Reactivity at Mineral-Water Interfaces Through Bulk Crystal Conduction**

*S. V. Yanina and K. M. Rosso*

A current flow through a hematite crystal couples dissolution and growth reactions at different surfaces, a finding likely relevant to a broad range of semiconducting minerals.

[10.1126/science.1154833](https://doi.org/10.1126/science.1154833)

### PHYSICS

**Frequency Ratio of Al<sup>+</sup> and Hg<sup>+</sup> Single-Ion Optical Clocks; Metrology at the 17th Decimal Place**

*T. Rosenband et al.*

Precise measurements of the frequency ratio of two optical clocks indicate that the fine-structure constant is fine and constant to an uncertainty of 10<sup>-17</sup>.

[10.1126/science.1154622](https://doi.org/10.1126/science.1154622)

### MEDICINE

**Oncogenic *CARD11* Mutations in Human Diffuse Large B Cell Lymphoma**  
*G. Lenz et al.*

One type of non-Hodgkin's lymphoma is caused by mutations in a scaffolding protein that inappropriately activate an inflammatory signaling pathway.

[10.1126/science.1153629](https://doi.org/10.1126/science.1153629)

### NEUROSCIENCE

**A Model for Neuronal Competition During Development**

*C. D. Deppmann et al.*

Modeling and experiments show that neurons successfully innervate muscle when neuronal sensitization to survival signals outweighs antagonistic signals for cell death.

[10.1126/science.1152677](https://doi.org/10.1126/science.1152677)

## TECHNICAL COMMENT ABSTRACTS

### GEOCHEMISTRY

**Comment on "Early Archaean Microorganisms Preferred Elemental Sulfur, Not Sulfate"** 1336

*H. Bao, T. Sun, I. Kohl, Y. Peng*

[full text at www.sciencemag.org/cgi/content/full/319/5868/1336b](http://www.sciencemag.org/cgi/content/full/319/5868/1336b)

**Response to Comment on "Early Archaean Microorganisms Preferred Elemental Sulfur, Not Sulfate"**

*P. Philippot et al.*

[full text at www.sciencemag.org/cgi/content/full/319/5868/1336c](http://www.sciencemag.org/cgi/content/full/319/5868/1336c)

## REVIEW

### MEDICINE

**An Oncogene-Induced DNA Damage Model for Cancer Development** 1352

*T. D. Halazonetis, V. G. Gorgoulis, J. Bartek*



1343 & 1377

## BREVIA

### ECOLOGY

**Dimethylsulfoniopropionate as a Foraging Cue for Reef Fishes** 1356

*J. L. DeBose, S. C. Lema, G. A. Nevitt*

Reef fish locate desirable feeding opportunities by recognizing dimethylsulfoniopropionate, which is released by coral reef algae in response to foraging by fish.

## RESEARCH ARTICLES

### CLIMATE CHANGE

**Long-Term Sea-Level Fluctuations Driven by Ocean Basin Dynamics** 1357

*R. D. Müller, M. Sdrolias, C. Gaina, B. Steinberger, C. Heine*

Considering changes in the shape of the ocean basins implies that global sea level fell by ~250 meters since 140 to 80 million years ago, a larger drop than implied in a recent study.

### PSYCHOLOGY

**Antisocial Punishment Across Societies** 1362

*B. Herrmann, C. Thöni, S. Gächter*

Retaliation against those who enforce the social norms of civic cooperation and rule of law varies among cultures and is more pronounced where social norms are weaker.

>> [Perspective p. 1345](#)

## REPORTS

### PHYSICS

**Fiber-Optical Analog of the Event Horizon** 1367

*T. G. Philbin et al.*

An optical analog of the event horizon of a gravitational black hole can be produced with light pulses propagating along an optical fiber, providing a tractable experimental system.

>> [News story p. 1321](#)

[CONTENTS continued >>](#)

## REPORTS CONTINUED...

### MATERIALS SCIENCE

**Stimuli-Responsive Polymer Nanocomposites Inspired by the Sea Cucumber Dermis** 1370

*J. R. Capadona et al.*

A rubbery polymer coated with nanofibers that segregate upon addition of a solvent can rapidly stiffen and relax, like the inner skin of a sea cucumber.

### CHEMISTRY

**Heterogeneous Nucleation Experiments Bridging the Scale from Molecular Ion Clusters to Nanoparticles** 1374

*P. M. Winkler et al.*

Experiments mimicking aerosol formation show that organic vapors tend to condense on any nanometer-scale particles, particularly negatively charged ones, rather than forming pure nuclei.

### GEOLOGY

**Age and Evolution of the Grand Canyon Revealed by U-Pb Dating of Water Table-Type Speleothems** 1377

*V. Polyak, C. Hill, Y. Asmerom*

Dating of cave deposits that form at the water table implies that incision of the Grand Canyon began in the west 17 million years ago and only accelerated in the east recently. >> *Perspective p. 1343*

### PLANETARY SCIENCE

**The Dust Halo of Saturn's Largest Icy Moon, Rhea** 1380

*G. H. Jones et al.*

Observations of a large shadow in electrons and particles behind Rhea imply that it has a small ring system formed from dust lofted into orbit by impacts.

>> *News story p. 1325*

### BOTANY

**TOPLESS Mediates Auxin-Dependent Transcriptional Repression During *Arabidopsis* Embryogenesis** 1384

*H. Szemenyei, M. Hannon, J. A. Long*

A transcriptional co-repressor is part of the protein complex that inhibits developmental gene activation in *Arabidopsis* until the growth hormone auxin triggers its degradation.

### BIOCHEMISTRY

**De Novo Computational Design of Retro-Aldol Enzymes** 1387

*L. Jiang et al.*

A computationally designed enzyme acts as a retro-aldolase that splits a carbon-carbon bond in a nonnatural substrate.

### MICROBIOLOGY

**A Cholesterol Biosynthesis Inhibitor Blocks *Staphylococcus aureus* Virulence** 1391

*C.-I. Liu et al.*

A drug for controlling cholesterol may be useful as an antibiotic for multi-drug-resistant *Staphylococcus* because of unexpected structural similarities among critical proteins.

### GENETICS

**High-Resolution Mapping of Crossovers Reveals Extensive Variation in Fine-Scale Recombination Patterns Among Humans** 1395

*G. Coop, X. Wen, C. Ober, J. K. Pritchard, M. Przeworski*

High-density genotyping of individuals from 82 families shows unexpected variation in the number of meiotic crossovers and in the relative activity of recombination hotspots.

### GENETICS

**Sequence Variants in the *RNF212* Gene Associate with Genome-Wide Recombination Rate** 1398

*A. Kong et al.*

A variant of a human gene associated with high rates of recombination in males and low rates in females is an ortholog of a nematode gene essential for recombination.

### CELL SIGNALING

**Hepatic Glucose Sensing via the CREB Coactivator CRTC2** 1402

*R. Dentin, S. Hedrick, J. Xie, J. Yates III, M. Montminy*

Competing glycosylation and phosphorylation of a single amino acid in a transcriptional coactivator regulate nutrient- and energy-sensing pathways and may contribute to diabetes. >> *Perspective p. 1348*

### MICROBIOLOGY

**Coiled-Coil Irregularities and Instabilities in Group A *Streptococcus* M1 Are Required for Virulence** 1405

*C. McNamara et al.*

Mutating a cell-surface virulence protein on strep bacteria stabilizes its structure, minimizing its inflammatory side effects and potentially making it a better vaccine.



SCIENCE (ISSN 0036-8075) is published weekly on Friday, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals Mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2008 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership and subscription (51 issues): \$144 (\$74 allocated to subscription). Domestic institutional subscription (51 issues): \$770; Foreign postage extra: Mexico, Caribbean (surface mail) \$55; other countries (air assist delivery) \$85. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request. GST #1254 88122. Publications Mail Agreement Number 1069624. SCIENCE is printed on 30 percent post-consumer recycled paper. Printed in the U.S.A.

Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. Postmaster: Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090-6178. Single-copy sales: \$10.00 current issue, \$15.00 back issue prepaid includes surface postage; bulk rates on request. Authorization to photocopy material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that \$20.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for Science is 0036-8075. Science is indexed in the Reader's Guide to Periodical Literature and in several specialized indexes.

CONTENTS continued >>



A child with progeria.

## SCIENCE NOW

[www.sciencenow.org](http://www.sciencenow.org) DAILY NEWS COVERAGE

### Too Old, Too Fast

Stem cell identity crisis might explain "accelerated aging" disease.

### How Weeds Take to Cities

A clever experiment shows one way to adapt to life in the concrete jungle.

### Physicists Successfully Store and Retrieve Nothing

Researchers trap a peculiar state of the vacuum.



Professional science master's grads are in high demand.

## SCIENCE CAREERS

[www.sciencecareers.org](http://www.sciencecareers.org) CAREER RESOURCES FOR SCIENTISTS

### Mastering the Job Market?

*B. Benderly*

Graduates of professional science master's programs often get multiple job offers.

### Getting Women Scientists to the Top

*C. Wald*

Programs in Europe aim to ensure that excellent junior-level women become senior-level scientists.

### Your Career in a Number

*G. Bjorn*

Measures of research impact are gaining popularity, but they have limitations with early-career faculty.

### March 2008 Funding News

*J. Fernández*

Learn about the latest in research funding, scholarships, fellowships, and internships.



Mechanical tension causes TGF- $\beta$  release.

## SCIENCE SIGNALING

[www.stke.org](http://www.stke.org) THE SIGNAL TRANSDUCTION KNOWLEDGE ENVIRONMENT

### PERSPECTIVE: Matrix Elasticity, Cytoskeletal Tension, and TGF- $\beta$ : The Insoluble and Soluble Meet

*R. G. Wells and D. E. Discher*

Extracellular matrix stiffness regulates the equilibrium between storage and release of matrix-bound transforming growth factor- $\beta$ .

### GLOSSARY

Find out what MVE, SFRP, and LCR mean in the world of cell signaling.

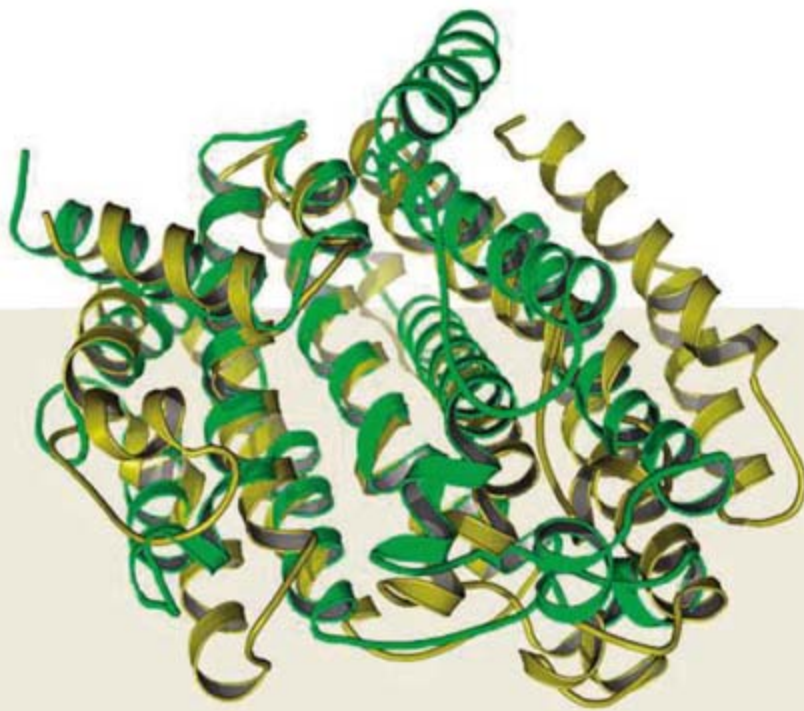


## SCIENCE PODCAST

Download the 7 March *Science* Podcast to hear about new insights on long-term sea-level change, widespread antisocial punishment, asteroids and comets looming near Earth, and more.

[www.sciencemag.org/about/podcast.dtl](http://www.sciencemag.org/about/podcast.dtl)

Separate individual or institutional subscriptions to these products may be required for full-text access.



## TOWARD DESIGNER ANTIBIOTICS

In the past few years, there has been an alarming increase in the number of methicillin-resistant *Staphylococcus aureus* (MRSA) infections occurring in hospitals and in the community. Liu *et al.* (p. 1391, published online 14 February) suggest an antibiotic approach that targets the pigment staphyloxanthin that, besides giving *S. aureus* its golden color, provides resistance to killing by the host immune system. Early steps in staphyloxanthin biosynthesis resemble those in cholesterol biosynthesis and an enzyme involved in pigment biosynthesis, CrtM (dehydrosqualene synthase), is structurally similar to human squalene synthase (SQS). Inhibitors of SQS were screened for activity against CrtM, and the structures of CrtM bound to three compounds that inhibited pigment formation in vitro were examined. One of these inhibitors, which has been through initial human clinical trials as a cholesterol-lowering agent, increased the susceptibility of *S. aureus* to killing by the innate immune system.

## Reconciling Sea-Level Changes

Estimates of global sea level during the Cretaceous period vary widely and sometimes are difficult to reconcile with the geological record of widespread continental flooding that occurred then. In order to remove some of those contradictions, Müller *et al.* (p. 1357) incorporate marine geophysical data into reconstructions of ocean basin volumes to show that global sea level was between 175 and 235 meters higher during the Cretaceous, when climate was much warmer than it is today. Moreover, by using a mantle convection model to incorporate the effect of the dynamic topography of the East Coast of North America associated with the movement of the subducted Fallarón plate, some different sea-level reconstructions can be reconciled that had seemed to be mutually exclusive.

## Rings Around Rhea?

Rhea, one of Saturn's largest moons, lies outside of the main rings of the planet and is heavily

cratered. As described by Jones *et al.* (p. 1380; see the news story by Kerr), Cassini recently flew by Rhea and was able to detect a depletion in ionized particles and electrons in Saturn's magnetotail downstream of Rhea out to several radii of the moon. Analysis of the data and the geometry of the depletion imply that Rhea may have a ring system formed from dust lofted from the moon by impacts that caused the observed depletions.

## Model Citizens?

Much has been made of the prosocial force of altruistic punishment in human societies—the willingness of some individuals to incur personal costs in order to coerce others into behaving in a cooperative fashion. Herrmann *et al.* (p. 1362; see the cover and the Perspective by Gintis) describe a large data set collected from university students across a range of societies. The frequency and extent of behaviors that were antisocial in nature—where individuals who had been punished would react by punishing the cooperators—was correlated with previous survey data measuring the social norms of civic cooperation and rule

of law in these same societies. Cross-societal variation of this antisocial punishment was associated with cross-societal variation in cooperative behavior and was more prevalent in societies with weaker social norms.

## Grand Canyon Dates

The incision history of the Grand Canyon has been an unsettled issue, in part because common dating methods that rely on the analysis of basalt flows and travertine do not provide any information on samples older than about 1 million years. In order to overcome this limitation, Polyak *et al.* (p. 1377; see the Perspective by Atkinson and Leeder) take advantage of technical advances in uranium-lead dating methods to date cave mammillaries, a type of speleothem that forms only in caves at or near the depth of the water table. Because these structures are common throughout the canyon, the authors could construct a history of its incision that extends back 17 million years for much of the length of the canyon.

## Changing Stiffness with Solvents

The sea cucumber can rapidly and reversibly alter the stiffness of its inner dermis. Capadona *et al.* (p. 1370) created a material that mimics these properties by combining a rubber copolymer and cellulose nanofibers taken from tunicates. Addition



of a hydrogen-bonding solvent causes the fibers to segregate. This process can be reversed upon removal of the solvent, which causes a reaggregation of the

fibers to form a percolating network. By interfering with the cellulose network in this way, the stiffness of the composite can be altered by a factor of 40.

## Seeing Over the Optical Event Horizon

The event horizon of a gravitational black hole represents a point of no return—particles inside this boundary, including photons, cannot escape. However, the direct observation of the

*Continued on page 1305*

Continued from page 1303

event horizon and the possibility of verifying the proposed theoretical properties appear remote. Looking to lab-based analogies that may describe the underlying physics, **Philbin *et al.*** (p. 1367; see the news story by **Cho**) report to have found a connection between light propagation in optical fibers and black hole physics. They report on the observation of an optical event horizon and probe some of the expected properties, such as frequency shifting of probe light. The authors also propose a scenario for observing Hawking radiation.

## Getting the Balance Wrong

Cancer development lies in the balance between tumor-suppressing activities and tumor-promoting activities. **Halazonetis *et al.*** (p. 1352) review the accumulating evidence that oncogenes, which force cells into rapid cycles of cell division that lead to cancer, cause errors in DNA replication and accumulation of damaged DNA. Cells that detect problems with DNA replication can signal through the p53 tumor suppressor protein and delay cell cycle progression. Mutations in p53 that are very common in cancer cells presumably bypass this barrier to tumorigenesis.

## Toward Designer Enzymes

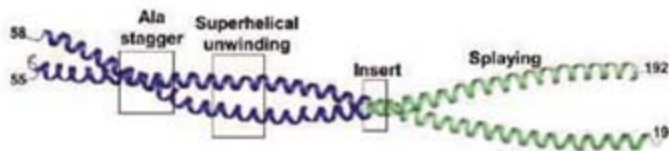
The creation of "designer enzymes" with new activities remains a challenge. **Jiang *et al.*** (p. 1387) take a step in this direction by computationally designing a retroaldolase that, in a multistep reaction, catalyzes the breaking of a carbon-carbon bond in a nonnatural substrate. Seventy-two designs representing four different catalytic motifs and 10 different scaffolds were examined. Active-site designs that used charged side-chain networks to mediate proton transfer were less successful than simpler designs that used a water molecule. Close agreement between two crystal structures and the corresponding design models validated the design procedure; however, the catalytic proficiency of the designed proteins remains far from that of natural enzymes.

## Toward a Designer GAS Vaccine

M proteins are antigenically variable cell-surface virulence factors of group A *Streptococcus* (GAS). This organism is an important cause of human illness and notably provokes the autoimmune disease rheumatic fever. **McNamara *et al.*** (p. 1405) examined a fragment of M1 that binds to host fibrinogen and promotes inflammation.

The M1 protein consists of  $\alpha$ -helical coiled coils bearing distinctive structural features that impart irregularity and instability to the coiled coil. These features mimic

muscle myosin and tropomyosin, which may explain postinfection autoimmune responses. Mice immunized with a mutant, stabilized M protein exhibited reduced harmful inflammation without compromising protective immune responses.



## Fine-Scale Analysis of Human Variation

Two reports examine patterns of human recombination that underlie some of the significant variation that exists among humans. **Coop *et al.*** (p. 1395, published online 31 January) genotyped individuals in a Hutterite population and identified the fine scale of recombination events between generations. **Kong *et al.*** (p. 1398, published online 31 January) examined parents in the Icelandic population to identify a gene associated with recombination frequency. Together, these studies point to a genomic region affecting the variation in recombination levels among individuals.

## Directly Modulating Glucose Homeostasis

The concentration of glucose in the bloodstream is regulated by glucose itself along with the hormones insulin and glucagon. Glucagon stimulates gluconeogenesis in part by regulating phosphorylation of a transcriptional coactivator known as cyclic adenosine monophosphate response element-binding protein 2 (CREB). **Dentin *et al.*** (p. 1402; see the Perspective by **Birnbaum**) found that high concentrations of circulating glucose also regulate CREB, but do so through stimulation of the hexosamine biosynthetic pathway and consequent O-linked glycosylation of the same serine residue in CREB that is modified by phosphorylation.

CREDIT: MCNAMARA ET AL.

# THE TACONIC ADVANTAGE

## ADME-Tox

Another Taconic breakthrough:  
The first and only portfolio  
of commercially available, fully  
licensed transgenic ADME-Tox  
mouse models.

Introducing:

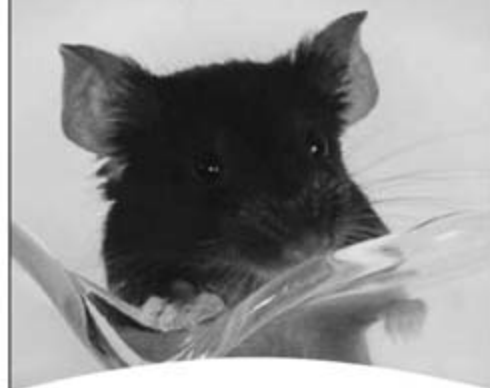
## THE HRN™ MOUSE (Hepatic Reductase Null)

Lacks hepatic cytochrome  
p450 activity

Evaluate efficacy of drug  
leads with small amounts  
of compound

Determine whether efficacy  
or toxicity is due to parent  
or metabolite

Available exclusively  
from Taconic.



Call today to put the Taconic  
Advantage to work for you.

# Taconic

Smart Solutions To Improve Human Health

1.866.515.2656

[www.taconic.com/HRN3](http://www.taconic.com/HRN3)



Bruce Alberts is  
Editor-in-Chief of *Science*.

## A New Editor-in-Chief

IT IS WITH RECOGNITION OF UNDERTAKING A HUGE NEW RESPONSIBILITY THAT I WRITE MY first Editorial as Editor-in-Chief of *Science*. New science will be needed on our crowded planet to protect our environment, to insure our health, and to provide sufficient water and food for humanity. This publication is a major vehicle for the success of the scientific enterprise, both by spreading scientific findings and by promoting community standards.

My predecessor as Editor-in-Chief, the former president of Stanford University, Donald Kennedy, has done a masterful job of steering this ship to make it maximally effective on both scores. His nearly 8 years of leadership have created a new emphasis on the science of sustainability, on science education, and on standards of openness and honesty in science. He has exposed our readers to advances in the social sciences, as well as in a wide range of the natural sciences. He has broadened our international reach, establishing news bureaus in Germany, South Africa, and China, adding to those in the United Kingdom, France, Japan, and India.

Don's strong support of young scientists is reflected in the *Science* Careers Web site that he promoted. Never afraid of controversy, he has written more than 150 Editorials, many of them boldly defending the scientific point of view on everything from genetically modified crops to climate change. He has repeatedly insisted on the highest standards of scientific conduct. Urbane, witty, and engaging, Don has set a nearly impossible standard to follow. I salute his past achievements, which leave this magazine in excellent shape.

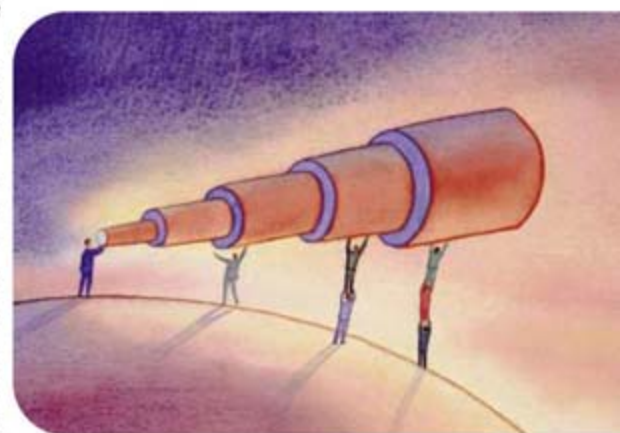
Why did I accept this position? In many ways I see it as an extension of my 12 years as president of the U.S. National Academy of Sciences (NAS). As part of the National Academies, the NAS published more than 200 reports each year. Most were in response to government requests, on topics that ranged from the health effects of arsenic in drinking water to the mentoring of science graduate students and postdoctoral fellows. I soon became painfully aware of the many opportunities to spread science and scientific ways of thinking that are being missed—in our failure to teach science as inquiry to most students, in our overly narrow definition of scientific careers in universities, and in the inadequate recognition of the truly international nature of science.

I also became aware that science is a remarkable social construct, which relies on constant vigilance to maintain the standards that make it so successful. It matters a great deal how we scientists treat our colleagues, reward excellence, share our data and resources, and relate to the public, whose taxes support the enterprise with generous research funding. Partha Dasgupta, Distinguished Professor of Economics at Cambridge University, has noted that "Today, we take it for granted that the Institution of Science has in place incentives which encourage researchers to disclose their findings for public use. But the emergence of those social contrivances which embody those incentives was not inevitable, nor did they emerge easily: It required the collective efforts of scientists and their patrons to establish them . . . the Institution of Science embodies a set of cultural values in need of constant protection . . ."

*Science* plays an important part in protecting this set of cultural values. It does so through the way that it selects the articles it publishes, enforces standards of honesty and data sharing, and makes all scientific articles freely available on the Web within a year after publication. Equally important, it produces a vigorous News section that attempts to probe every aspect of science and its interface with society with truth and integrity.

I am at heart a scientist, and scientists love to tackle important challenges. The challenge of this new job is to vigorously explore how I and *Science's* outstanding staff might make this publication an even more effective agent for spreading science and its values throughout the world—a world that desperately needs much more of both.

— Bruce Alberts







## ECOLOGY

## A Flashy Affair

Humans see chameleons as masters of camouflage. The lizards themselves seem to see more in it than predator avoidance. This family of lizards has evolved a range of talents, some merely changing skin tone, others changing color, too. All thanks to rapidly reacting neurally controlled chromatophores in their skin. Stuart-Fox and Moussalli used reflectance spectroradiometry, tuned to the spectral sensitivities of the retinas of chameleons and of potential bird predators, to measure the conspicuousness of a variety of lineages of African dwarf chameleons in their preferred habitats. If the habitat is dense and complex, bird predators tend to be excluded, and there is less need for disguise. Here, the chameleons indulge in flashy aggressive behavior, changing color spectacularly in social displays; at least from a chameleon's-eye view. In more open, hotter habitats it pays to live a drab and quiet life, avoiding the notice of predators, staying cool sand-colored and merely adjusting brightness during social exchanges. So, deep in the bush, what we perceive as subtly changing camouflage, chameleons perceive as flamboyant social signaling. — CA

*PLoS Biol.* 6, e25, 10.1371/journal.pbio.0060025 (2008).

## CLIMATE SCIENCE

## All Stirred Up

Around 55 million years ago, at the height of the Paleocene-Eocene thermal maximum, the world was a much warmer place than today. Sea surface temperatures were higher everywhere than now, and the equator-to-pole thermal gradient was much shallower. Climate for much of the past 500 million years has been warmer than it is now, and during the warm periods the surface meridional temperature gradient generally appears to have been weak. Explaining how the climate system might have transferred heat from low to high latitudes to maintain such a shallow thermal gradient has been difficult, and many hypotheses have been advanced, including those involving effects from radiative forcing by high concentrations of atmospheric carbon dioxide, more intense ocean heat transport, differences in the amounts and locations of polar stratospheric clouds, and extra-tropical atmospheric convection. Koryt *et al.*, using a coupled model of intermediate complexity, investigate another possibility: that tropical cyclones could have caused enough ocean surface mixing in the tropics to cool the sea surface there and drive the

strong poleward heat flux needed to produce the shallow thermal gradients that seem to have prevailed during warm climates. This solution, if correct, also has implications for how we might expect the climate system to respond to anthropogenic warming. — HJS

*J. Clim.* 21, 638 (2008).

## BIOCHEMISTRY

## Anyone for a Cuppa?

Inhaling the soothing aroma from a mug of peppermint tea renews the sense of wonder at the delightful mix of monoterpenes produced by the hybrid plant *Mentha* × *piperita*. Its oil, which

accumulates within a structure called a peltate glandular trichome, contains high levels of (–)-menthol and low levels of the precursor (+)-pulegone and the side product (+)-menthofuran. The critical balance between the activities of pulegone reductase (which produces menthones that are converted into menthols) and mentho-

furan synthase (which diverts pulegone into menthofuran) can be disturbed in less-than-ideal environments, such as low light.

Rios-Esteva *et al.* have collected the measured biochemical parameters of kinetic constants and metabolite concentrations, validated them where feasible against independently derived quantities (gland volume, oil density, and terpene molecular weight), and built a biosynthetic model that simulates the developmental time courses under optimal and low-light conditions of menthol, pulegone, and menthofuran levels. An earlier genetic engineering study had suggested that menthofuran turned down the transcription of pulegone reductase, but plugging this constraint into the model failed to reproduce the observed changes in terpene levels. Instead, the observations could be replicated by assuming that menthofuran acted as a competitive inhibitor of the reductase, a supposition confirmed by Lineweaver-Burk analysis of *in vitro* enzyme assays. — GJC

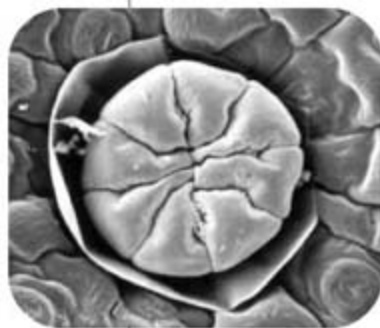
*Proc. Nat. Acad. Sci. U.S.A.* 105, 2818 (2008).

## CHEMISTRY

## A Fruitful Fuel Proposal

Sustainable alternatives to gasoline and diesel fuels will need to recapitulate some of the properties that make currently used fuels attractive

*Continued on page 1311*



Scanning electron micrograph showing the peltate glandular trichomes.

Continued from page 1309

energy carriers, such as low vapor pressure and stable long-term storage. Horváth *et al.* examined the properties of one such alternative,  $\gamma$ -valerolactone (GVL)—a small cyclic ester produced by fruits and used as a food additive. Compared to methanol, ethanol, and methyl and ethyl *tert*-butyl ethers, GVL has a much lower vapor pressure, which reduces volatile emissions, and also low melting and flash points that facilitate safe storage. It is not readily hydrolyzed to the acid under pH-neutral conditions, nor does it readily form peroxides. Because GVL does not form an azeotrope with water, less energy is needed to recover it through distillation. The authors have explored catalytic routes to GVL from sucrose and conclude that with further advances in synthetic efficiency from biomass, GVL could serve as an oxygenated fuel additive or even a fossil-fuel replacement on its own. — PDS

*Green Chem.* **10**, 238 (2008).

## APPLIED PHYSICS

## Vapors Plucked on a Harp

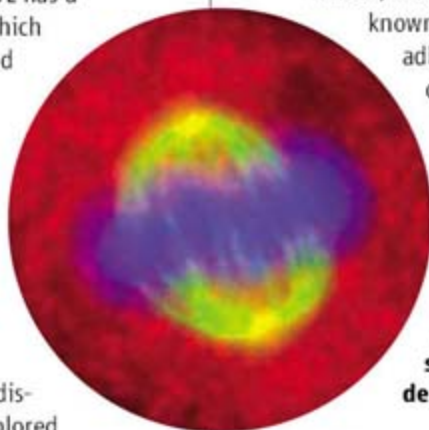
Optical or electronic devices that sniff out chemical compounds with high sensitivity and selectivity clearly have applications in a broad range of circumstances, from environmental monitoring to the detection of biohazards. However, the compounds of interest may not be in pure form; often they are diluted among multiple background substances, and such an artificial nose must therefore be able to sort through complex mixtures. Stievater *et al.* present a chemical sensor based on a micro-optomechanical bridge that can be interrogated remotely using optics. Their device, a microharp, forms one end of a Fabry-Perot interferometer and comprises an array of microbridges in which each bridge is coated with a sorbent polymer sensitive to a particular analyte. Each bridge also has a distinct length and so vibrates at an individual frequency. When the device is placed in a test atmosphere, different compounds bind to particular strings on the harp, changing the mass of the bridge and thereby inducing a shift in the frequency at which it vibrates. Interrogating the vibrational frequency of each bridge separately allows monitoring of a vapor's chemical composition remotely, and in some cases (as for organophosphonates) detection concentrations down to 17 parts per billion. — ISO

*Opt. Express* **16**, 2423 (2008).

## CELL BIOLOGY

## Doing Double Duty

During cell division, centrosomes are important in the organization and maintenance of the mitotic spindle. Fielding *et al.* found that a kinase, integrin-linked kinase (ILK), known to be important in cell adhesion, also interacts with centrosomal and spindle proteins, including the main constituent of the mitotic spindle, tubulin. When ILK was inhibited, cells



**ILK (red) concentrates at the centrosome at the poles of the mitotic spindle (green). Condensed chromosomes (blue).**

failed to assemble mitotic spindles properly, resulting in aberrant chromosome segregation. The centrosomal protein RUVBL1 was important for ILK targeting to the centrosome, and the activity of ILK was important for another centrosomal protein, ch-TOG, to promote spindle pole organization and mitosis. This unanticipated role for ILK in promoting centrosomal organization, spindle assembly, and chromosome segregation may represent an important link between cell adhesion and mitosis. — SMH

*J. Cell Biol.* **180**, 681 (2008).

## HUMAN GENETICS

## An Autism Association

Although autism is highly heritable, sorting out the genes associated with this complex disease has been difficult. Weiss *et al.* searched for structural mutations (duplications or deletions below the level of microscopic detection) in the genomic DNA of 751 families who are part of the Autism Genetic Resource Exchange. They found a significant association of autism with a nearly 600-kb region that was deleted or duplicated at a locus on chromosome 16. This structural mutation was also observed in patients from Children's Hospital in Boston and in a group from Iceland. It occurred at a frequency of approximately 1% in patients as compared with less than 0.1% of the general population. It might result from unequal crossing over at this region of chromosome 16, which is a known hotspot for deletion and duplication and is bordered by two duplicated regions. Autism may represent the summation of a series of rare events, whose detection will require screening tens of thousands of patient samples. — BJ

*N. Engl. J. Med.* **358**, 667 (2008).



Order today,  
oligos tomorrow

Next-Day Service for  
custom DNA oligos

Invitrogen's new U.S.  
express service allows you to order  
oligos and receive them the next  
business day. Just order online  
before 2 p.m. eastern time.

Get 2 OD units guaranteed minimum yield  
(25 nmol scale) for tube DNA oligos 7 to 40  
bases in length, with the same quality control  
as our standard oligos.

On-time shipment is guaranteed—if we  
don't ship your full order on time, the \$19.95  
Next-Day Service fee will be credited back  
to you. Next-Day Service is available in the  
United States only.

Place your order now at  
[www.invitrogen.com/oligos](http://www.invitrogen.com/oligos).

 invitrogen™

[www.invitrogen.com](http://www.invitrogen.com)

©2008 Invitrogen Corporation. All rights reserved. Service for U.S. customers only.  
These products may be covered by one or more Limited Use Label Licenses (see  
Invitrogen catalog or [www.invitrogen.com](http://www.invitrogen.com)).



## Budding Scientists

How does your garden grow? A team of ecologists and climate scientists wants to know. Project BudBurst invites the public to record the timing of leaves, flowers, and fruits to help track climate change. Part of the USA National Phenology Network (NPN), the project began in 2007 to foster consistent continentwide monitoring of phenology—the timing of annual biological cycles. Participants from across the United States enter their reports on the Project BudBurst Web site, and university and government researchers will analyze the data.

Jake Weltzin, a government ecologist and NPN executive director, says the network aims to understand how climate variation affects ecosystems. Plant phenology is a “sensitive integrator” of environmental factors, he says. For example, the timing of lilac blooms in the western United States is one of the best ways to predict a wildfire year.



To monitor how plants respond to climate, NPN will maintain a database that includes contributions from trained researchers and Project BudBurst participants. In 2007, a test run of BudBurst fielded 913 observations from 26 states. Sandra Henderson, a science educator and Project BudBurst coordinator, says the program captured the public imagination because of the climate change connection. “People don’t want to sit on the sidelines and be passive observers,” she says.

## Without Bounds

Bound encyclopedias can’t hold a cursor to the multimedia display of the first part of the Encyclopedia of Life (EOL), a Web site assembled by a consortium of 25 museums, botanical gardens, and other scientific institutions. The effort, funded in part by \$12.5 million in grants, aims to provide one-stop shopping for biodiversity for scientists and the public. Launched last year (*Science*, 11 May 2007, p. 818), EOL will come online in stages. The first phase, unveiled in late February, includes information on 30,000 plants and

NET  
WATCH

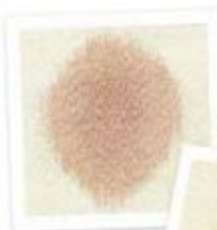
animals. Eventually, the site will incorporate images, distribution maps, life histories, and identification data for each of Earth’s 1.8 million known species.

The first entries were built using multiple existing sources, an effort the collaborators hope to automate in the future. The first set of pages concentrates on fish and amphibians, as well as plants related to peppers, tomatoes, and petunias. Work on the project is ongoing. However, scientists note, given the rapid rate of new species discovery, it’s a project that likely will never be finished.

>> [www.eol.org](http://www.eol.org)

## Let the Sun Shine

Researchers from Australia and Hong Kong plan to take the wash out of washday with fabrics that clean themselves. The self-cleaning property comes from coating the fibers with nanocrystals of titanium dioxide, a photocatalyst that decomposes dirt and stains when exposed to light. Walid Daoud, a chemist at Monash University in Churchill, Australia, and colleagues at the Hong



**New TiO<sub>2</sub>-infused fabric is self-cleaning when exposed to light.**



Kong Polytechnic University describe their findings in the 26 February issue of *Chemistry of Materials*.

The photocatalytic property of titanium dioxide is well known. Daoud and his colleagues developed a treatment for fabric fibers that chemically bonds the titanium dioxide nanocrystals to the fibers. Although sunlight produces the best results, the self-cleaning works under any light source, even while the clothes are being worn. The crystals also inhibit the growth of odor-causing bacteria.

Daoud foresees initial applications in clothing for campers and soldiers. “But the target is really to reduce the consumption of water, detergent, and energy used in laundering and dry cleaning,” he says. The treatment shouldn’t increase the cost of fabric production, says Daoud, who is negotiating with potential industrial partners. Self-cleaning fabrics could be on the market in 2 years.

## True Blue

Bowls, ritualistic objects—even people—were painted a brilliant turquoise blue before being offered up as sacrifices by the Maya, the original residents of the Yucatán Peninsula. Now, 14th century pottery recovered from an ancient sacrificial well reveals just how the Maya created the blue pigment.

The Maya started using the pigment known as “Maya blue” about 500 C.E. Researchers knew that its ingredients included indigo and the clay mineral palygorskite, but the rest of the preparation process was unknown. Studies of a Maya bowl found at the bottom of the Sacred Cenote, a famous sacrificial well in Chichén Itzá, showed traces of a resin incense called copal. Researchers now believe that when heated, the copal fused the indigo and palygorskite together, producing the long-lasting blue pigment.

Archaeologist Dean Arnold of Wheaton College in Wheaton, Illinois, came across the bowl, which had originally been uncovered in 1904, while examining collections at the Field Museum in Chicago. In Maya culture, each of the three ingredients had healing power, says the Field Museum’s Gary Feinman. “The Postclassic Maya appear to have used this pigment as a key component in rituals that petitioned for rain to heal the Earth from drought and desiccation,” he says. Arnold’s team reported the findings online 26 February in the journal *Antiquity*.





**HEROIC.** Michael Kelley, an oncologist and basic researcher at the Veterans Affairs Medical Center in Durham, North Carolina, routinely gets calls from cancer patients asking him for health advice. But Josh Sommer is the first patient he met whose reaction to being diagnosed was to “roll up [his] sleeves and say, ‘I want to work with you and in your lab.’”

Sommer, a junior at Duke University in Durham, found out in 2006 that he had chordoma, a rare cancer that afflicts one in a million people with tumors at different spots along their spinal column. The news prompted him and his mother, physician Simone Sommer, to launch the Chordoma Foundation to foster collaborations among the small community of chordoma researchers around the world. Sommer, who switched from environmental engineering to biomedical engineering after learning of his illness, also works part-time for Kelley analyzing the gene expression of different chordoma and normal cell lines.

“I think they’ve catapulted the research light-years ahead,” says Duke’s Neil Spector, a cancer drug development expert who has found preliminary evidence that an existing chemotherapy drug may prove helpful against chordoma. Sommer is “a phenomenally gifted young man,” Spector adds.

## THEY SAID IT

“It is a moral outrage that a wealthy country like the United States allows its closest neighbors to suffer from some of the world’s worst levels of disease, poverty, and malnutrition. ... By transforming Gitmo from a detainee facility to a center for research on the diseases of poverty, the U.S. would show that it sincerely wants to address the Millennium Development Goals in Latin America and the Caribbean, and ultimately make things better for the next generation of all Americans.”

—Peter Hotez, editor-in-chief of *PLoS Neglected Tropical Diseases*, offering a bit of unsolicited foreign policy advice about Guantánamo Bay in the February issue of the journal.

## MOVERS

**A KEEPER.** Six months after a stalled attempt to appoint a head of geosciences, the National Science Foundation (NSF) has named space physicist Timothy Killeen to the job. Killeen currently heads the National Center for Atmospheric Research (NCAR) in Boulder, Colorado.

The geosciences directorate has been



without an assistant director since January of last year. NSF asked oceanographer Mark Abbott of Oregon State University (OSU), Corvallis, in July of last year to fill the job. Abbott later withdrew after NSF’s lawyers told him that his close managerial ties to OSU posed too many conflicts of interest (*Science*, 17 August 2007, p. 879). NCAR receives 70% of its \$150 million budget from NSF, but Killeen says he’s stepping down as director to become a senior scientist and “will not be involved with funding

decisions, advocacy, and other matters relative to NCAR.”

Killeen says his 8 years as director of NCAR have taught him that being a science administrator “is all about people, facilities, managing effectively, and keeping an eye on the science.” He’ll join NSF in July after completing his 2-year term as president of the American Geophysical Union in Washington, D.C.

Got a tip for this page? E-mail [people@aaas.org](mailto:people@aaas.org)

## In the News >>

**A SECOND LIFE.** Korean veterinarian Byeong Chun Lee, who collaborated with disgraced stem cell researcher Woo Suk Hwang to create the world’s first cloned dog, Snuppy, is now trying to clone a deceased pit bull terrier named Booger.

Booger was owned by a California woman, Bernann McKinney, who banked some of his ear tissue after his death in 2000. She recently placed a \$150,000 order with a Seoul-based biotech firm, RNL Bio, to have the dog cloned. Under an agreement between RNL Bio and Seoul National University, the cloning will be done by Lee and his colleagues at SNU. If successful, it will be the first commercial cloning of a dog in the world, which could launch an era of pet cloning. (The tissue bank company McKinney used, Genetic Savings and Cloning, successfully cloned a cat for \$50,000 in 2004 before being acquired by Viagen, a California company that clones livestock.)

Lee made headlines in 2005 as part of a research team that cloned an Afghan hound. Although most of Hwang’s other accomplishments turned out to be fraudulent, Snuppy stood up to scientific scrutiny. During its investigation of the Hwang affair, SNU officials found Lee guilty of embezzling funds but allowed him to stay on because of his accomplishments.

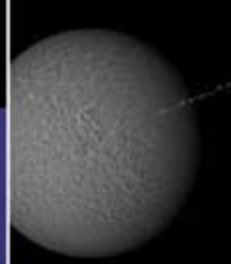
Since Snuppy, Lee has cloned a number of other dogs and a wolf. He plans to deliver Booger’s clone by February 2009.





Costing out big science at NSF

1322



A ring around a Saturn moon?

1325

## NATURAL DISASTERS

# Ecologists Report Huge Storm Losses in China's Forests

**GUANGZHOU, CHINA**—From delicate orchids and magnolias to rare Chinese yews and Kwangtung pines, the flora of Guangdong Nanling National Nature Reserve is considered so precious that ecologists call the reserve “a treasure trove of species.” But winter storms have reduced the biological hot spot to a splintered ruin. Snow, sleet, and ice laid waste to 90% of the 58,000-hectare reserve's forests, says He Kejun, director of the Guangdong Forestry Administration in Guangzhou.

Nanling Reserve is one of scores of fragile ecosystems, from Anhui Province in the east to Guangdong Province in the south, that took a beating from storms in late January and early February that set records for snowfall and low temperatures in some areas. Last week, China's State Forestry Administration (SFA) announced that the storms damaged 20.86 million hectares—one-tenth of China's forests and plantations—roughly equivalent to the number of hectares that were reforested between 2003 and 2006. SFA pegs the losses at \$8 billion. “The severe storms did a massive amount of harm,” says Li Jianqiang, a plant taxonomist at Wuhan Botanical Garden. “This scale of damage has never happened before.” He Kejun and others say it will take decades for the hardest-hit ecosystems to recover.

The ecological and economic toll rivals that of devastating floods along the Yangtze River in 1998 that inundated 25 million hectares of farmland. For broadleaf evergreen forests, “this is bigger than the Yangtze disaster. It's unique in the history of south China,” says Ren Hai, an ecologist with the South China Botanical Garden (SCBG) in Guangzhou. SFA and other agencies have dispatched scientists to take stock and formulate restoration plans. “The government is acting very, very fast,” says Ren.

In southeastern China's worst winter in 5 decades, snow and ice knocked out power and paralyzed roads and rail lines at the height

of the year's busiest travel season—the Spring Festival, when many Chinese return to their hometowns. The storms pummeled 21 of 33 provinces and regions, claiming 129 lives. Some 485,000 homes were destroyed and another 1.6 million damaged,



**Disaster scene.** Storm damage was more severe in eight provinces (red); devastation at Nanling reserve.

displacing nearly 1.7 million people, according to central government statistics. Agriculture officials estimate that 69 million livestock—mostly chickens and ducks—froze to death. Storm-related losses exceed \$21 billion. As *Science* went to press, electricity had still not been restored to some remote areas.

Scenes of scums at train stations and vehicles adrift on highways were splashed across the news in China and abroad last month. Meanwhile, outside the spotlight, an ecological calamity was unfolding. In Jiangxi Province, for example, entire bamboo forests were reduced to matchsticks; fast-growing bamboo can regenerate in several years. In Guangdong, officials estimate that more than

700,000 hectares of forest and plantations are damaged severely, with losses approaching \$1 billion. Other provinces enduring extensive forest damage are Anhui, Guangxi, Guizhou, Hubei, Hunan, and Sichuan (see map).

The carnage was not limited to natural ecosystems. “Exotic species were harmed more than native species,” says Ren. In northern Guangdong Province, plantations of slash pine (*Pinus elliottii*), an import from the southern United States, splintered under wet snow, and extensive stands of Australian gum trees “are almost all going to die,” Ren predicts. At Wuhan Botanical Garden in Hubei Province, the roof of a greenhouse housing Asia's largest assemblage of aquatic plants caved in under heavy snow. “A unique collection has been lost,” says Wuhan botanist Li Xiaodong.

SCBG scientists maintain long-term experimental plots at Nanling that will allow them to gauge ecosystem damage and recovery. At the moment, the picture is bleak. Nanling's entire forest between 500 meters and 1300 meters in elevation was wiped out, says He. “Before the storm, we could hear birds singing in the reserve. Now it is mostly silent,” he says. Many *bai xian*, or silver pheasants—Guangdong's official bird—succumbed to the severe weather, and carcasses litter Nanling's trails, says He. One worry, he says, is that epidemics will erupt this spring in the storm-sapped animal populations and among migratory birds.

With support from Guangdong Province's government, SCBG plans to send teams of scientists to several of the most devastated forests to survey damage and to set up test plots that will track everything from species composition to the susceptibility of the degraded forests to insect pests and fires.

The storm damage lends urgency to a new national strategy for plant conservation released last week by SFA, the Chinese Academy of Sciences, and the State Environmental Protection Agency. Under the manifesto, crafted with help from Botanic Gardens Conservation International, a Richmond, U.K., nonprofit, China has pledged to launch a nationwide survey of species and habitats, construct a national herbarium, crack down on illegal logging, and establish by 2010 a system to monitor and protect China's 31,000 plant species, more than half of which are native. Some ▶



5000 plant species in China are threatened with extinction.

As damage assessments proceed, SFA has established a disaster relief technology group and will hold an emergency meeting later this month to plan for restoration. Botanical gardens are doing their part, too. "We must work hard to save vegetation and lessen the extent of damage," says Ren. "We want to find a way to help natural ecosystems

recover with minimal human disturbance."

That is a tricky balancing act. At Nanling, managers are barring local residents from entering to remove downed timber. Although salvage logging could reduce wildfire risk, it could exacerbate erosion, further degrading ecosystems. The bulk of the restoration work is likely to focus on economic recovery: rehabilitation of plantations. The storm's aftermath should also

spur long-term research on plant cold tolerance, says Li Jianqiang.

The immediate task is picking up the pieces after the worst winter in recent memory. "We cherish our endangered species," says Li. But for some of the precious plants at Wuhan Botanical Garden and in southern China's battered reserves, he says, "there is nothing we can do to save them."

—RICHARD STONE

With reporting by Li Jiao in Beijing.

## ANIMAL RIGHTS

# Brazilian Scientists Battle Animal Experimentation Bans

Brazilian scientists are fighting a series of local attempts to ban animal experimentation that they say could cripple scientific research. At the top of their list: a controversial law passed 2 months ago by the city of Rio de Janeiro that prohibits all animal experiments at private companies. Researchers are hoping that a comprehensive federal bill addressing animal experimentation, which has been lingering in the Brazilian Congress for 12 years, will put a stop to such local bans.

The battle in Rio, a major biomedical research hub, has gone on for more than 2 years. In 2006, the city council passed an animal-welfare law, introduced by actor-turned-politician Cláudio Cavalcanti, that would have banned all animal experiments in the city. Mayor César Maia vetoed the bill. A second version, passed in September 2007, made an exception for universities and public organizations such as the Instituto Oswaldo Cruz (Fiocruz), a major vaccine producer. Maia vetoed that one, too. But the council overrode his veto on 26 December.

The law has not taken effect yet, and the mayor does not appear in a hurry to enforce it, says animal physiologist Luis Eugênio Mello of the Federal University in São Paulo, president of the Federation of Brazilian Societies of Experimental Biology. But if enforced, the ban could force several Rio biotech companies out of business. "It's a crazy law," says Eduardo Krieger, a former president of the Brazilian Academy of Sciences.

A comparable far-reaching bill was approved in December by Florianópolis, the capital of the southern state of Santa Catarina; that law was replaced by the city's mayor in February with much less

stringent regulations. Similar legislative plans are afoot in other cities.

Brazil's scientists contend that regulating animal research should not be a local issue and are arguing for a federal law. Indeed, such a bill was introduced in 1995 by Chamber of Deputies member Sérgio Arouca, who was once director of Fiocruz; it would ban animal experiments if other alternatives are available, require ethics committees to approve studies, and set up a national council to issue guidelines. But the bill never came to a vote, and Arouca died in 2003.

Researchers say "Arouca's law" would protect them from a wave of municipal or state initiatives, and they have been lobbying hard to get the bill to a vote. Animal-rights activists oppose it, however. Ethics panels, which already exist at the majority of research institutions, are dominated by scientists and rubber-stamp proposals, says George Guimarães, director of Ethical Vegetarianism, Animals Rights Defense and Society, a São Paulo-based group.

The researchers' lobbying appears to have paid off, says Mello, with "support from left to right" in Congress. And recently, Brazil Presi-



**No petty issue.** Cláudio Cavalcanti is pushing for a ban on animal experiments to include the Instituto Oswaldo Cruz (top).

dent Luiz Inácio Lula da Silva, who has made advancing research a national priority, named Arouca's law among his legislative priorities. Guimarães agrees that the federal bill is now likely to pass. But a legislative stalemate in the Congress, unrelated to the bill, could make it hard to pass any laws at all in 2008, Mello warns.

Cavalcanti says that he wants Rio's mayor to enforce his law; he will also reintroduce the proposal for a total ban this year. Officials at Fiocruz, a big yellow fever vaccine producer, have warned that such a move could imperil routine quality testing of vaccines. But Cavalcanti says that he does not believe animal research can benefit human health. A single-issue politician, he has offered to become a guinea pig himself if it can help save animals. "This is my mission, my only reason for living," he says.

Scientists should do more to counter the cruel image of animal studies that activists have promoted and explain why such work is necessary, says Walter Colli of the University of São Paulo. "We are guilty of not having done enough to influence public opinion," he says. "The average citizen is confused."

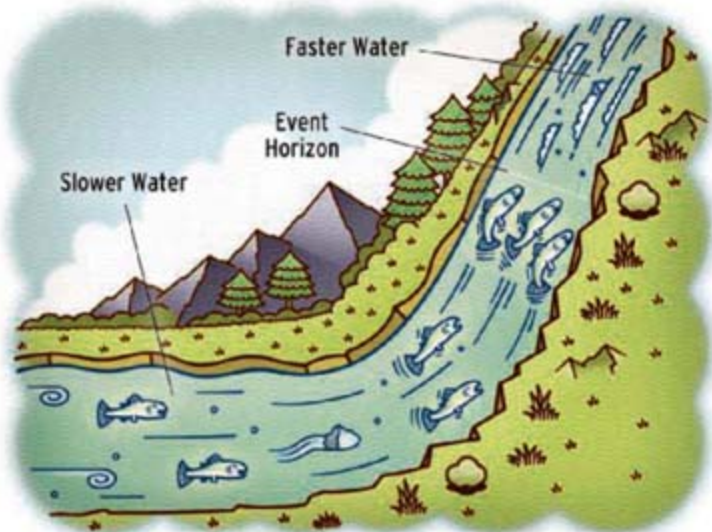
—MARTIN ENSERINK

## PHYSICS

# Test of Hawking's Prediction on the Horizon With Mock 'White Hole'

Physicists can't travel to a black hole to see how it ticks, but they have taken a big step toward creating something similar in the lab. Using an optical fiber and laser light, Ulf Leonhardt of the University of St. Andrews in the U.K. and colleagues have simulated a "white hole"—essentially a black hole working in reverse—as they report on page 1367. The model might soon mimic perhaps the most tantalizing property of a black hole: the "Hawking radiation" that should emanate from it.

Others have cooked up analogs, but "this is probably the first one that has some correlation to the Hawking effect," says Grigori Volovik, a physicist at Helsinki University of Technology



**No go.** Current can stop fish moving upstream and mimic an event horizon. A pulse in an optical fiber captures the physics, too.

in Finland, who is working on a model in liquid helium. "This is some kind of a breakthrough."

Formed from the collapse of a star, a black hole is like a funnel-shaped pit in the fabric of space and time. Light or anything else that ventures into the funnel cannot get back out once it passes a point of no return known as the event horizon. In contrast, a white hole would resemble a mountain in spacetime so steep that nothing could reach the summit. It would have an event horizon that marks the point of closest approach. Unlike black holes, white holes should be unstable, and none are thought to exist.

In principle, simulating an event horizon is simple. Consider a river filled with fish swimming upstream at maximum speed. If the fish reach a point where the water flows as fast as they can swim against it, they will pile up there (see figure, above). That point simulates

a white-hole horizon, and theorist William Unruh of the University of British Columbia in Vancouver, Canada, and colleagues have experimented with fluid analogs.

Leonhardt and colleagues took a different tack. They fired a pulse of infrared light down an optical fiber. Crucially, the intensity of the pulse itself altered the speed at which light could travel in the fiber. The researchers then shot in light of a second, longer infrared wavelength that moved slightly faster than the pulse. As that light caught up to and "ascended" the pulse, it slowed until at some point its speed exactly matched that of the pulse.

That spot on the pulse simulated the event horizon, and light accumulating there was compressed to a slightly shorter wavelength. That squeeze made the light travel slower than the pulse, so it effectively rolled back down the peak and fell behind. The researcher detected the telltale wavelength shift and other evidence that the light was piling up on the horizon.

The next goal is to see Hawking radiation, Leonhardt says. Thanks to quantum mechanics, the vacuum roils with photon pairs that normally pop in and out of existence too quickly to be observed. But near a black hole, one particle might emerge beyond the

event horizon and fall in while its partner emerges outside and escapes, as theorist Stephen Hawking argued in 1975. So a black hole ought to glow like an ember, although far too feebly to be seen through the cosmic microwave background.

The white-hole analog ought to radiate, too, Leonhardt says. Like warped spacetime near a black hole, the wildly varying speed of light in the fiber can rip particles out of the vacuum. So the pulse should shine faintly but detectably in ultraviolet wavelengths.

Seeing that glow would be key, Unruh says, because in spite of its fame, Hawking radiation remains unconfirmed. If physicists spot its equivalent in an analog, Unruh says, "then you get a lot, lot more faith that the prediction is solid." Volovik agrees: "If they really see Hawking radiation, I think Hawking will finally get his Nobel Prize." —ADRIAN CHO

## Crash in Antarctica Kills Two

A helicopter crash in Antarctica has claimed the lives of two people and injured three. Willem Polman, 45, a technician at the Royal Netherlands Institute for Sea Research (NIOZ), was killed on 2 March when a helicopter based on the research ship *Polarstern* crashed near the German Antarctic station Neumayer II. The German pilot, 37, was also killed. Three other passengers were injured, two of them seriously, according to the Alfred Wegener Institute (AWI) in Bremerhaven, Germany, which operates both the ship and the coastal Neumayer station.

The ship is on a 10-week voyage to investigate the Southern Ocean as part of the International Polar Year. One of the injured is Dutch scientist Maarten Klunder, 27. Two others, a 24-year-old German helicopter technician and a 25-year-old female French researcher, were also injured. NIOZ announced the names of the two Dutch researchers, but AWI has declined to name the pilot and the other researchers, citing German privacy laws. The news of Polman's death "hit all of us like a bomb," says Jan Boon, a NIOZ spokesperson. "We still have trouble believing it." After the injured are evacuated, *Polarstern* is expected to continue its voyage, which is scheduled to end in Punta Arenas, Chile, on 16 April. The cause of the crash is unclear; the weather at the time was apparently good. AWI is working with the German Aviation Authority to investigate.

—GRETCHEN VOGEL AND MARTIN ENSERINK

## New Suit for Penguins

The U.S. government is moving too slowly to protect 10 species of penguins against climate change, says a suit filed last week by the Center for Biological Diversity (CBD), a Tucson, Arizona-based environmental group. According to legal code, the U.S. Fish and Wildlife Service (FWS) was supposed to decide by the end of November whether to list the penguins under the Endangered Species Act, which would require tighter regulations on fisheries and review of greenhouse gas emissions by government agencies. Chris Tollefson, an FWS spokesperson, says a decision will be finalized in the "next few weeks."

But environmental physiologist Yvon Le Maho of the French national research agency in Strasbourg, France, says that without a "big movement against climate warming, [official protection of the penguins] may be hopeless." If the U.S. government's answer is no, CBD says it will sue again.

—ELSA YOUNGSTADT



## LARGE FACILITIES

# NSF Delays Three Projects to Get Better Handle on Costs

After a decade of making their case to the U.S. National Science Foundation (NSF), scientists planning a major project for remote monitoring of the oceans thought they had cleared the final hurdle in December. That's when an external panel blessed the \$331 million venture, called the Ocean Observatories Initiative (OOI), and told NSF officials "to enter into the detailed design and construction phase" to build it. "We were ready to go, and the reviewers agreed," says Steven Bohlen of the Consortium for Ocean Leadership in Washington, D.C., which is managing the project.

So Bohlen and his colleagues were shocked last month when NSF omitted building funds for OOI and two other long-running projects on the verge of construction—the \$100 million National Ecological Observatory Network (NEON) and the \$123 million Alaska Region Research Vessel (ARRV)—from its 2009 budget request to Congress. It's part of a new policy aimed at eliminating cost overruns that occur after construction is under way. Those overruns have not only forced NSF to borrow from other accounts, but they can also lead to last-minute changes that weaken a project's scientific capabilities. Under the previous policy, a project was approved based chiefly on its scientific merit; it might be years before NSF arrived at a final price based on all relevant factors. Now, NSF is requiring a firm cost estimate before asking Congress for construction funds.

Scientists whose projects face delays of a year or more aren't pleased about the sudden policy shift. But they understand why NSF is asking them to go back to their calculators. "It's not good news for the science," says Terry Whitlege, director of the Institute of Marine Science at the University of Alaska, Fairbanks, which manages the ARRV project for NSF. "But I think that NSF is probably doing the right thing in the current budget environment. And their message to all of us is clear: Don't come back to us

down the road with a higher number."

The new rules mark the latest attempt by NSF, an agency known for its expertise in small science, to get a better handle on an expanding portfolio of large projects. The facilities can cost upward of \$500 million to build and tens of millions of dollars a year to operate, a sizable commitment for a \$6 billion agency. The policy affects four projects now in NSF's major research equipment and facilities construction account—including the Advanced Technology Solar Telescope in Hawaii, which is listed as a new start for 2009 although it is currently still in the design phase—as well as all future proposals.

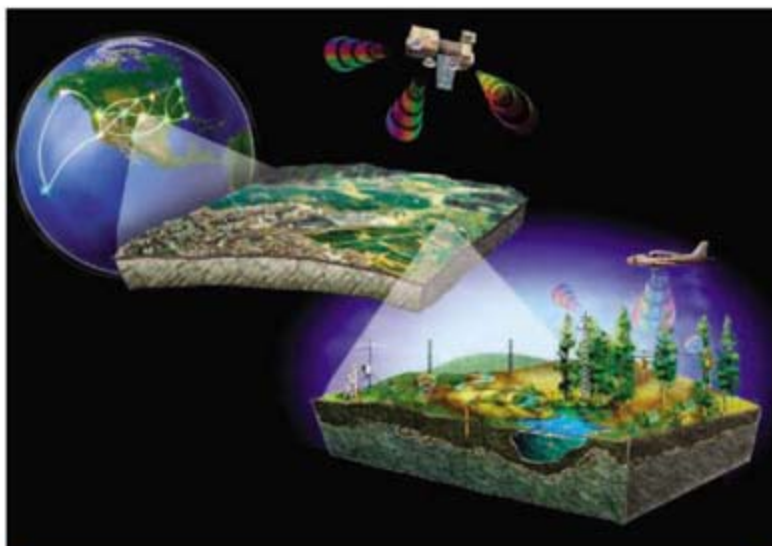
The large-facilities account was created more than a decade ago to segregate big-ticket items such as ships and telescopes from the agency's bread-and-butter research and education programs. But the community was so unhappy with how NSF's over-

sight body, the National Science Board, approved and ranked projects that it complained to Congress, which ordered up a review by the National Academies' National Research Council. Its 2004 report recommended that the process be more rigorous and transparent (*Science*, 16 January 2004, p. 299). NSF Director Arden Bement says the agency has embraced those suggestions by setting up a new administrative office and monitoring each project more closely.

Even so, last month Bement went one step further. Although NSF has spent millions on each project to help scientists lay the groundwork, Bement says he won't ask for construction funds until each has passed a final project review that includes a firm cost estimate and a detailed analysis of environmental and regulatory issues. "It's a huge culture change for the foundation," says a White House official familiar with how NSF manages its large projects.

A more thorough vetting could shorten the actual construction time, Bement says, and reduce the chances that a project would need to be "descoped" to stay within its budget. He says the new rule also brings NSF practices closer to those at other federal agencies, such as the Department of Energy and NASA, which have more experience building and managing large scientific facilities and instruments.

NEON's checkered history highlights the problems Bement is trying to correct. NSF first requested money for NEON—some \$12 million—in its 2001 budget request. Then—NSF Director Rita Colwell called it "a continental-scale research instrument consisting of 10 geographically distributed observatories, networked via state-of-the-art communications, for integrated studies to obtain a predictive understanding of the nation's environments." But the initial design was reworked substantially after scientists raised numerous objections. Last spring, the community came up with the current version, which features a network of 20 core



**From ship to shore.** A national ecological network (top) and a new Arctic research ship have been temporarily pulled from NSF's construction budget.

CREDITS (TOP TO BOTTOM): NICOLLE RAGER FULLER/NSF; GLOSTEN ASSOCIATES, PROVIDED BY SPOCS, UNIVERSITY OF ALASKA, FAIRBANKS

sites and 40 “relocatable” sites. The core sites are expected to provide a 30-year longitudinal record of myriad factors, whereas the other sites will focus on narrower scientific questions and capture more transient environmental events.

Despite those ups and downs, its price tag never varied. David Schimel, who runs the Boulder, Colorado-based consortium responsible for building NEON, says the original \$100 million figure announced in 2001 “was not based on anything.” The new policy, he says, allows project leaders to do it right. “NSF has actually done us a huge favor by unshackling us from that \$100 million estimate,” he says. “Now we can start over and come up with a new, more realistic baseline. Needless to say, the new figure will be higher.”

Just as important as the initial construction cost, says Schimel, is the estimated \$30 million a year needed to operate and maintain the network. “That’s the real constraint,” he says. “We don’t want to gut the community’s research budget [at NSF] by building a facility that’s too costly to operate.” Project scientists are hoping to incorporate several features to reduce labor and maintenance costs in the final design, he says. Accordingly, NSF’s 2009 budget request includes \$26 million for NEON from its research account, in part to fund the additional work needed to come up with a more efficient design.

Project leaders for OOI, which is expected to cost \$50 million annually to operate, say they made similar hard choices in preparing for the preliminary design review NSF conducted in December. The observatories will gather data on coastal, regional, and global scales, and the community has been ruthless in paring each system down to the bone, says Holly Given, the consortium’s director of ocean-observing activities. For example, Bohlen notes that only three sites remain from an original plan for 10 blue-water autonomous buoy systems—in the Southern Ocean off the Chilean coast, the North Atlantic near Greenland, and the Gulf of Alaska in the northern Pacific. “As we refined our cost estimates, we had to scale back and concentrate on what was most important scientifically,” he explains.

The cost of some components can’t be nailed down until the plans are actually sent out for bids, Bohlen says. Referring to the five sets of seabed cables that will connect



**Oceans 3.** OOI has been scaled back to three blue-water autonomous buoy systems and two regional networks.

instruments continuously monitoring the Juan de Fuca Ridge off the coast of Washington state, he notes that “the market price for those materials and sensors, plus labor, can vary a lot.” NSF is seeking \$10.5 million in 2009 for OOI for continued planning.

Scientists involved in the Alaska research vessel are acutely aware of how the economy can wreak havoc on carefully laid scientific plans. Whitley estimates that NSF’s new policy will add 12 to 18 months to the project’s scheduled solicitation of bids in 2010—at a price yet to be calculated. “Shipyard costs have been going up by 20% a year,” Whitley says, because of the rising cost of steel and other raw materials and industry’s demand for new and refurbished exploration ships triggered by \$100-per-barrel oil prices. A delay also means a longer wait for data on the impacts of climate change in the Arctic, he notes. The ship will replace the *Alpha Helix*, a 40-year-old research vessel that the university retired in 2004.

Bement doesn’t pretend to have all the answers for managing large facilities. “I’m convinced that we can do a lot better,” he told Representative Alan Mollohan (D-WV), chair of the House panel that sets NSF’s budget, during a hearing last week on NSF’s 2009 budget request. But the problem clearly has his full attention. When Mollohan asked about one project, Bement brushed aside the chair’s suggestion that he turn to one of his aides for the details. That won’t be necessary, Bement replied: “I know the answer. I get a report every month. And I read them.”

—JEFFREY MERVIS

## Brits Rejoin Gemini

The United Kingdom last week reversed plans (*Science*, 23 November 2007, p. 1227) to withdraw from the Gemini Observatory, whose twin scopes sit in Hawaii and Chile. Instead, it will stay put and try to save some of the £4 million in annual costs it’s obligated to pay by selling telescope time. The plan to pull out of the international partnership outraged British astronomers because it threatened to cut off their access to the northern skies. “It’s unfortunate this was so badly handled,” says astrophysicist Martin Rees, president of the U.K.’s Royal Society.

—ELIZABETH QUILL

## Wisconsin Wins Latest Patent Bout

The U.S. Patent and Trademark Office (PTO) has upheld a 2006 patent on human stem cells owned by the university-affiliated Wisconsin Alumni Research Foundation (WARF). Two years ago, two nonprofits challenged the patent and two other WARF patents. Last week, the government released an 85-page decision that upholds the patents yet narrows their scope slightly. “We’re very pleased,” says WARF’s managing director, Carl Gulbrandsen. “We believed from the very beginning that [James] Thomson’s discoveries were patentable.” WARF’s opponents, including Alan Trounson, now president of the California Institute for Regenerative Medicine, had argued that Thomson’s work was obvious when he performed it in the 1990s. But the patent examiner rejected those arguments earlier this week, saying that the published science in the 1990s was too “unpredictable” to lead someone to try making human stem cells with an “expectation of success.” The groups plan to appeal the decision, and rulings on the other two patents are pending.

—ELI KINTISCH

## Research Strategy: Centered On Centers

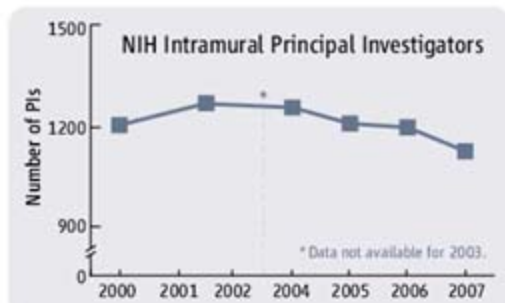
The U.S. Department of Homeland Security (DHS) has funded five new university-based research centers to study border security and immigration; explosives detection; maritime, island and port security; natural disasters; and emergency management and transportation. The new centers, each of which will receive up to \$2 million a year for 4 to 6 years, will join eight existing centers focused on DHS’s mission. “These colleges and universities ... will provide scientific expertise, high-quality resources, and independent thought—all valuable to securing America,” says Jay Cohen, DHS undersecretary for science and technology.

—YUDHIJIT BHATTACHARJEE

## RESEARCH FUNDING

## U.S. Biomedicine's Mother Ship Braces for Lab Closings

Distress signals are emerging from the intramural program at the U.S. National Institutes of Health (NIH) in Bethesda, Maryland, as funding troubles begin to pinch. Most institutes are affected, but the pain is acute at the National Institute of Child Health and Human Development (NICHD), where up to 12 intramural labs—run by 16% of 74 tenured staff—could be shuttered. “This is a completely new category of nightmare,” says an NICHD investigator who asked not to be named. Compared with a poor review in the extramural world, in which a researcher can try for a new grant, closing an intramural lab means going “from full funding to zero,” he says.



**Downsizing.** The NIH intramural program has lost more than 100 PIs since growth ended in 2003.

NICHD's troubles reflect the impact of 5 years of flat budgets on the \$2.8 billion NIH intramural program. The campus has seen a net loss of 114 of 1252 principal investigators (PIs), or 9%, since 2004 when a period of rapid growth halted. Half of the decline came in the past year, according to NIH data. “There’s no way with conservation of matter to do anything else,” says NIH Deputy Director for Intramural Research Michael Gottesman, who nevertheless thinks the program is still “a reasonable size.”

He adds that the squeeze “is not unique to NIH or any organization,” although extramural research seems less constrained. The number of funded extramural PIs has hovered around 26,300 for the past 4 years, according to NIH.

Gottesman points out that the intramural program has downsized before, after a 1994 blue-ribbon panel called on NIH to cut less productive programs and create a formal tenure system. The number of PIs dropped from roughly 1584 in 1990 to 1206 in 2000. Growth resumed from 2000 to 2002 (see graph). But when it stopped, many of NIH’s 21 intramural programs had a hard landing.



**Constant gardener.** Intramural research chief Michael Gottesman says there are benefits to pruning.

Scientific directors saw budgets lag behind inflation while costs increased.

As a result, at the National Cancer Institute, the tally of PIs has dropped by 65 to 253, a decline of about 20% since 2003, says Center for Cancer Research Director Robert Wiltout. The institute has been more aggressive in closing labs after a leader retires or receives low marks on a site visit. And some top scientists have simply left. The diabetes institute closed several labs in 2006 to help trim 7% from its operating budget, says National Institute of Diabetes and ▶

## SPACE PHYSICS

## Antimatter Experiment May Be Too Costly for NASA to Launch

NASA says it is willing to fly a \$1.5 billion experiment designed to detect antimatter. But Congress would have to come up with as much as \$4 billion to make it happen, the agency says. Supporters of the Alpha Magnetic Spectrometer (AMS) dispute those cost estimates but face an uphill struggle to get the 7000-kg probe into orbit.

In a 17-page report to Congress that was released last month, NASA paints a sobering picture of what it would take to attach the instrument to the international space station. Samuel Ting, the physics Nobel laureate at the Massachusetts Institute of Technology in Cambridge who has championed the project, says the 16-nation AMS collaboration has no money to buy another ride into space. That leaves the fate of AMS, and its quest to understand why there is more matter than antimatter in the universe, with Capitol Hill.

Testifying before the House Science and Technology Committee on 13 February, NASA Administrator Michael Griffin said

he doesn’t oppose launching AMS aboard the shuttle. But for planning purposes, he added, Congress must find the money this year. The agency’s report, submitted a week later, notes that all remaining shuttle flights are devoted to completing the station by 2010, after which NASA intends to retire the launcher. Adding an additional flight “would be difficult, costly, and would have a significant negative impact on NASA’s exploration program,” the report asserts.

An additional flight in late 2010 would cost between \$300 million and \$400 million, NASA estimates. The cost would rise 10-fold if the flight were delayed until 2011 because of the need to extend industry contracts. “These costs would come directly at the expense of exploration development activities,” states the study.

NASA is overstating the costs, says Trevor Kincaid, a spokesperson for Representative Nick Lampson (D-TX), who represents the area around NASA’s Johnson Space Center

and is a strong AMS backer. Lampson puts the price of another shuttle mission at between \$150 million and \$175 million. A spokesperson for Senator Bill Nelson (D-FL) suggested that a \$1 billion supplement to NASA’s 2009 budget could cover the additional mission. Last summer, the Senate approved such an increase, but it was dropped from the final 2008 spending bill. “NASA signed on to this, and they should uphold their end of the [international] deal,” adds Kincaid.

The only option besides an additional shuttle mission is to put AMS on an expendable rocket. That would cost at least \$600 million, says the report, and delay the launch by 2 or 3 years while the spacecraft is modified. That means AMS might not arrive until 2013 or 2014 at a space station that NASA intends to shut down in 2016.

Ting and his collaborators say AMS needs at least 3 years in space to gather data. But the Nobel laureate says he’s not giving up. “My job is to get this finished,” he says. —ANDREW LAWLER

CREDITS: (IMAGE) NIH RECORD; (DATA) NIH OFFICE OF INTRAMURAL RESEARCH

Digestive and Kidney Diseases senior scientist Alan Schechter. Last year, a shortfall in lab operating funds at NICHD forced scientists to curtail experiments and travel (*Science*, 18 May 2007, p. 968).

The possible cuts in personnel at NICHD, however, appear to be more drastic than any before. Soon after the final NIH budget passed Congress in mid-December, Owen Rennert, NICHD's scientific director, met with program chiefs and "tentatively outlined ... some areas that could be reduced" to free up \$15 million, he told *Science* by e-mail. In January, a few PIs were told their labs were to be closed this year. After staff protested to Gottesman, Rennert relented. In a 30 January e-mail sent to *Science* and circulated to his staff, he wrote that "no decisions have been reached." Any cuts in pro-

grams, he wrote, would be based on reviews by outside scientists over the next 2 years and factors such as publications and relevance to NICHD's mission. Still, anxiety is running high throughout NICHD's labs.

Some don't blame Rennert; he has made "Herculean efforts" to avoid lab cuts until now, says one senior scientist. But PIs who are due for their 4-year site visit in April are now bracing for the worst. Those whose labs are closed won't be out of a job, Gottesman says, but will have to join someone else's lab or become an extramural grants administrator. University positions seem out of the question for anyone but superstars. Even if they survive, NICHD scientists worry about the impact: "You'll have to be more focused and not take as many risks," says fruit fly geneticist Judith Kassis.

It seems unlikely that NIH's intramural researchers will get much sympathy from outside, where funding is also tight. Yale University cell biologist Barbara Ehrlich, a member of NICHD's board of scientific counselors, says that although some of its investigators are "just spectacular," others "haven't kept up as much."

And Gottesman says there's a bright side to this "pruning": It has freed up some money to strengthen other labs, including big teams doing cutting-edge science. (The overall number of intramural scientific staff, about 6000 to 7000 M.D.s and Ph.D.s, is probably stable, he says.) "For the remaining scientists, this is still a terrific place." He admits, however, that "everybody says that I'm a Pollyanna." **-JOCELYN KAISER**

## PLANETARY SCIENCE

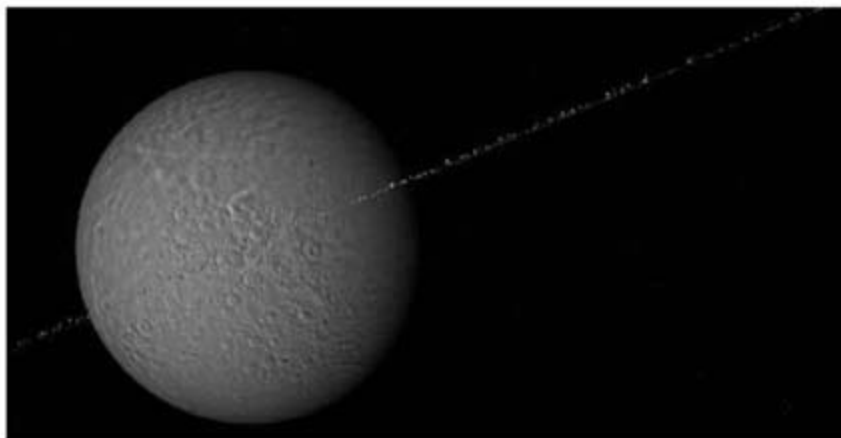
# Electron Shadow Hints at Invisible Rings Around a Moon

Space physicists poring over Cassini spacecraft data think they have two firsts: the first known natural satellites of a moon, which also form the first rings known to encircle a moon. Unlike the rings around Jupiter, Saturn, Uranus, and Neptune, however, the proposed rings around Saturn's moon Rhea are—so far, at least—invisible.

Even though the proposed rings are physically "very weird and [Cassini scientists] don't have image proof, it's certainly a good batch of circumstantial evidence," says ring dynamicist Jack Lissauer of NASA's Ames Research Center in Mountain View, California. Ring specialist Jeffrey Cuzzi, also of NASA Ames, agrees, up to a point. "It's clearly something unusual," he says. But whereas the discoverers "want to say it's unusual rings, I'd want to say it's unusual physics."

The case for rings depends on shadows Cassini passed through while flying by 1530-kilometer-diameter Rhea in November 2005, as space physicist Geraint Jones, now of University College London, and his colleagues report on page 1380. Rather than recording how the rings block starlight—a common way of spotting them—Cassini serendipitously recorded the dimming of Saturn's trapped energetic electrons that stream by Rhea. The electron dimming extended about 6000 kilometers on each side of Rhea.

Something seemed to be absorbing electrons before they got to Cassini. Because other Cassini instruments failed to detect enough gas or dust to do the job, Jones and colleagues inferred that unseen boulders up to about a meter across were absorbing the electrons.



**Now you see it.** In an artist's simulation, the proposed rings of Rhea—presented here edge-on—are obvious. In reality, there is no trace of them in spacecraft images, only in the dimming of flowing electrons they apparently absorb.

The Pioneer 11 spacecraft discovered the G ring of Saturn in 1979 in much the same way.

The clincher for Jones was a set of six dark, narrow electron shadows that Cassini recorded, three on each side of Rhea. The shadows are stunningly symmetrical side to side, as seen in figure 4b of the paper. "I saw [a figure] like that in my first planetary science class," Lissauer recalls. The year was 1977, and the figure was the now-classic plot of the previously unimagined narrow rings of Uranus occulting a star. Given Cassini's broad

electron absorption with narrow spikes, all strikingly symmetrical, a broad disk of debris in Rhea's equatorial plane with three embedded narrow ringlets or incomplete ringlets was "the only reasonable explanation we've been able to come up with," Jones says. Space physicist Mihaly Horanyi of the University of Colorado, Boulder, agrees. Cassini has provided "a fascinating set of observations from multiple instruments that is indeed best explained by proposing a possible set of rings around Rhea," he writes in an e-mail.

But ring specialists still have their reservations. Such rings, they say, are possible but improbable. First, just the right sort of impact would probably have been required to blast material off the icy moon and into orbit. Then the ring particles would have had to survive millions if not billions of years being torn apart by the tidal pull of Saturn and worn down to dust by eroding small impacts. Most constraining, perhaps, is the "incredibly low" limit on dust around Rhea set by Cassini's camera, says Joseph Burns of Cornell University, who is on the imaging team. Ring boulders must shed some dust, and even tiny amounts of dust show up when backlit by the sun. "We're going to keep trying harder," says Burns, to see what may be seen. **-RICHARD A. KERR**

# Preparing for Doomsday

**Over the next several years, new telescopes will spot thousands of near-Earth asteroids and comets. If one is headed our way, will world leaders be ready to respond?**

**TIESHAN TEMPLE NATIONAL FOREST, CHINA—**In the control room of XuYi Observatory, Zhao Haibin sits at a computer and loads the night sky over Jiangsu Province. A faint white dot streaks across a backdrop of pulsating stars. “That’s a satellite,” Zhao says. Elsewhere on the screen, a larger white dot lumbers from east to west. It’s a main-belt asteroid, circling the sun between Mars and Jupiter.

On a ridge in this quiet, dark corner of southeastern China, about 100 kilometers northwest of Nanjing, XuYi’s new 1-meter telescope spies a few dozen asteroids on a good night. Most are known to science. But since China’s first telescope dedicated to asteroid detection saw first light early last year, Zhao’s team has discovered more than 300 asteroids, including a near-Earth object (NEO), the class of asteroids and comets that could smash into our planet, if fate would have it.

China’s asteroid hunters are the latest participants in a painstaking global effort to catalog NEOs. Close encounters with asteroids in recent years—and comet Shoemaker-Levy’s spectacular death plunge into Jupiter in 1994—have spurred efforts to find the riskiest NEOs before they blindsides us. Tracking potentially hazardous objects—NEOs pass-



Hit or miss? Artist's conception of Apophis approaching Earth.

ing within 0.05 astronomical units, or 7.5 million kilometers, of Earth’s orbit—is essential for any attempt to deflect an incoming rock.

The first test of our planet’s defenses could be Apophis, an asteroid the size of a sports arena that made the world sweat for a few days in December 2004, when calculations suggested as great as a 1 in 37 chance of an impact in 2029. Although further data ruled out that day of reckoning, another could be looming. In April 2029, Apophis will pass a mere 36,350 kilometers from Earth, inside the

orbits of geostationary satellites. If it enters a keyhole—a corridor of space barely wider than the asteroid itself where gravitational forces would give it a tug—it will end up on a trajectory that would assure a collision 7 years later: on 13 April 2036, Easter Sunday. The odds of Apophis threading the needle are currently 1 in 45,000—but dozens of factors influence asteroid orbits. Researchers will get a better look during Apophis’s next appearance in our neighborhood in 2012.

By then, a powerful new telescope for detecting asteroids and comets—the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS), expected to be up and running by summer—should have unmasked thousands more NEOs. An even grander project, the 8.4-meter Large Synoptic Survey Telescope (LSST), is expected to be operational in 2014.

The anticipated bumper crop of NEOs confronts society with urgent questions. In the next several years, with increasing rapidity, Pan-STARRS and its ilk will discover potentially dangerous NEOs. Currently, 168 NEOs have a chance of striking Earth in the next century, although the odds are minuscule. By 2018, the risky rock roster could swell more than 100-fold. Additional observations will allow astronomers to refine orbits, and in most cases, rule out a threat. For that reason, astronomers are debating when the public should be alerted to hazards, to minimize false alarms.

**Postcard from the edge.** A tour guide puts the scale of Arizona’s Meteor Crater in perspective.

Eventually, an asteroid with our name on it will come into focus, forcing an unprecedented decision: whether to risk an interdiction effort. “The very concept of being able to slightly alter the workings of the cosmos to enhance the survival of life on Earth is staggeringly bold,” says Russell Schweickart, chair of the B612 Foundation, a Sonoma, California, nonprofit that lobbies for NEO deflection strategies. We have the means to deflect an asteroid—indeed, “it’s really the only natural hazard that we can possibly prevent,” says NEO specialist David Morrison, an astrobiologist at NASA’s Ames Research Center in Mountain View, California.

There is one “fatal missing element,” says Schweickart, who in 1969 piloted the lunar module for the Apollo 9 mission: “There is no agency in the world charged with protecting the Earth against NEO impacts.” He and others hope to change that.

### Wake-up calls

Like any natural disaster, impacts occur periodically; gargantuan impacts are so rare that their frequency is hard to fathom. Every 100 million years or so, an asteroid or a comet a few kilometers or more in width—a titan like the rock thought to have wiped out the dinosaurs 65 million years ago—smacks Earth. “This is not just getting hit and killed,” says Edward Lu, a former astronaut who now works for Google. “You’re on the other side of the Earth and the atmosphere turns 500° hotter. Lights out.”

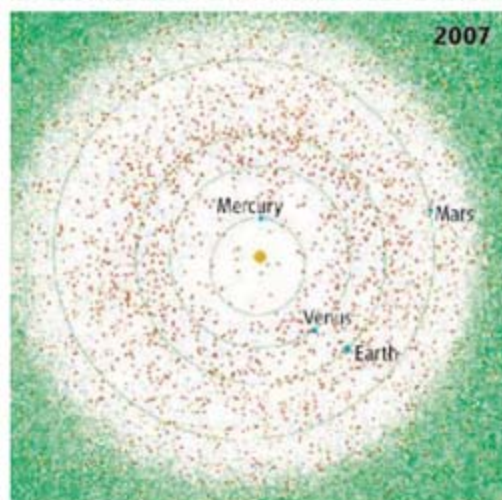
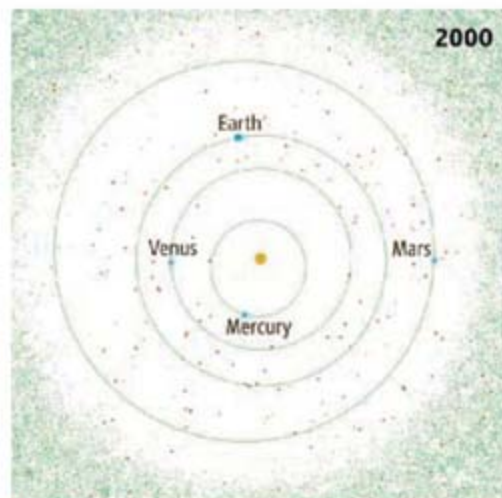
Reassuringly, no doomsday asteroid identified thus far is on track to intersect Earth’s orbit in the next century. Less reassuring, an unobserved, long-period comet from the Oort cloud could swoop in with little warning. Although the odds of this happening in anyone’s lifetime are on the order of winning the Powerball lottery, a megaimpact’s annualized fatality rate is likely to rival those of earthquakes or tsunamis, says Clark Chapman, an astronomer at the Southwest Research Institute in Boulder, Colorado.

Near-Earth asteroids tens to hundreds of meters in diameter are far more numerous—there may be as many as 3 million in the solar system—and they cross Earth’s path more frequently. The iconic Meteor Crater in northern Arizona was gouged by a 50-meter-wide hunk of iron and nickel 50,000 years ago. In 1908, a fireball scorched and flattened trees over 2100 square kilometers of taiga in Siberia’s Tunguska region—the devastating footprint, many experts say, of a modest asteroid that exploded in midair.

Recent supercomputer modeling has downsized the Tunguska rock. An asteroid just a few dozen meters wide, fragmenting

explosively with a yield of 3 to 5 megatons—a fraction of earlier estimates—could have done the trick, Mark Boslough and David Crawford of Sandia National Laboratories in Albuquerque, New Mexico, report in an article in press in the *International Journal of Impact Engineering*. If this is correct, the expected frequency of Tunguska-sized impacts changes from once every couple of millennia to once every couple of centuries. “Smaller objects may do more damage than we used to think,” says Chapman.

Today the impact threat may seem obvious,



**Aswarm with asteroids.** In 2000, there were more than 86,000 known asteroids. By 2007, there were nearly 380,000, including main-belt objects that don’t approach Earth (green); objects that approach but do not cross Earth’s orbit (yellow); and objects that cross Earth’s orbit (red).

but for decades it was largely ignored. Aerodynamicist Anatoly Zaitsev, director general of the Planetary Defense Center in Moscow, sounded the alarm in a landmark report delivered to Soviet leaders in 1986. “They just laughed,” he says. Then on 22 March 1989, an asteroid several hundred meters across whizzed by Earth at about twice the distance to the moon; astronomers didn’t spot Asclepius until it had already passed.

Asclepius was a shot across the bow, prompting the U.S. Congress to query NASA about whether the agency had a plan for the next killer asteroid. A parade of committees followed, after which Congress in 1998 ordered NASA to tally and track at least 90% of NEOs that are more than 1 kilometer wide. NASA launched the Spaceguard Survey, named after a survey in Arthur C. Clarke’s 1972 novel *Rendezvous with Rama*. To date, Spaceguard and other efforts have identified more than 700 of an estimated 1000 or so NEOs in this category. Then in 2005, Congress called on NASA to expand the search by 2020 to cover 90% of NEOs at least 140 meters in diameter—the approximate minimum size to damage an area at least as large as a state or seaboard. NASA expects Spaceguard II to spot 21,000 potentially hazardous NEOs and forecasts a 1-in-100 chance that such a rock will hit Earth in the next 50 years.

The uncertainties are huge. Main-belt asteroids can knock into each other, turning a benign rock into a malignant projectile. And with only a fraction of NEOs having been identified so far, what we don’t know can hurt us. Astronomer Brian Marsden, director emeritus of the International Astronomical Union’s Minor Planet Center, the clearinghouse for asteroid and comet orbits, figuratively sums up the situation: “The ones to worry about are those that were discovered yesterday and have a very high probability of hitting us the day after tomorrow. Those, plus the ones we’ve never even seen yet!”

### Drawing a bead

Night has fallen on an early December evening near Tieshan Temple, which, according to local lore, was the home of China’s first monk. The sky above the national forest is pitch-black but overcast. On nights like this, asteroid hunters know how to kill time. In a chilly, cigarette smoke-filled lounge down the hall from XuYi’s control room, Zhao and his colleagues play cards and sip from tall, clear plastic bottles packed with green tea leaves, hoping that the weather forecast is wrong and the skies will clear.

Zhao has worked at Purple Mountain Observatory, which operates XuYi, since graduating from Nanjing University in 1996. He has a comet named after him, but his biggest thrill came last spring, when he found an NEO.

On most nights, the telescope is pointed away from the sun, toward main-belt asteroids outside Earth’s orbit. More elusive objects between Earth and the sun can be discerned in the right conditions. With a clear sky and a new moon, just after nightfall or before sunrise, Zhao aims the telescope at a 60° angle to the

sun, where faint NEOs, like a crescent or gibbous moon, reflect sunlight in phases. During the telescope's first year, his team got fewer than a dozen opportunities to gaze sunward. One was 7 May, when they scored their NEO.

Tonight, just after midnight, the clouds have dispersed enough for viewing. Zhao's team swings into action, pointing the telescope at a 2-degree-square patch of sky. As dawn breaks, they will e-mail the data to Purple Mountain's Nanjing headquarters for analysis.

Zhao's team is working fast to stake NEO claims before Pan-STARRS, the first Spaceguard II facility, starts gobbling up the heavens. The telescope on Mount Haleakala on Maui Island, Hawaii, has a charge-coupled device camera with 1.4 billion pixels—the highest resolution in the world—that acquires images every 30 seconds.

Pan-STARRS, which saw first light last August, will usher in a new paradigm in observational astronomy (*Science*, 12 May 2006, p. 840). "It's a set of surveys that will be analyzed in a wealth of different ways," says Kenneth Chambers, an astronomer with the Institute for Astronomy (IfA) at the University of Hawaii, Manoa, who is leading a consortium of 300 scientists whose institutions have paid for first crack at Pan-STARRS gold. Some will map the Milky Way or look for distant quasars. Others will hunt for asteroids. "The astronomical community is not ready for the fire hose of data that's going to hit them," Chambers says.

Once Pan-STARRS begins taking data in earnest this summer, NEO finds should come thick and fast. According to IfA astronomer Robert Jedicke, who led development of the software that will cull NEOs from the data deluge, Pan-STARRS will be 10 times more effective at spotting NEOs than all current surveys combined. "Are there many more objects like Apophis out there? This is something

**In the hunt.** Zhao Haibin's team at XuYi has a few hundred asteroids under its belt.



that Pan-STARRS will answer," says IfA Director Rolf-Peter Kudritzki.

Magnificent feats of detection are also expected from LSST, which will have 24 times greater survey power than Pan-STARRS. Like its Hawaiian rival, the \$389 million project has broad science objectives, including studying dark energy and dark matter and mapping the Milky Way. Unlike Pan-STARRS, LSST data will be available immediately to any researcher. Construction is expected to begin in 2011 at Cerro Pachón, Chile.

When completed, LSST will cover the entire available sky every 4 nights with a 3.2-billion-pixel camera. Project scientists have teamed up with Google, Microsoft, and others to develop algorithms for processing the masses of data. After 10 years of operation, LSST should have plotted rough orbits for 82% of potentially hazardous NEOs larger than 140 meters, with only the risk assessments requiring human input, says LSST Director J. Anthony Tyson, a physicist at the University of California, Davis.

Funding is not assured. Tyson has lined up \$45 million so far from private sources, including two gifts announced in January that will help pay for the mirror: \$20 million from Charles Simonyi, chief executive of Intentional Software, and \$10 million from Microsoft's Bill Gates. The tycoons, says Tyson, "are excited about the LSST acting as a peripheral device for the Internet and thus bringing the universe to everyone's computer." Much of LSST's construction funds are expected from the U.S. National Science Foundation, which will hold a Major Research Equipment and Facilities Construction review on the project this autumn.

### Gauging risks

In the early 1990s, as astronomers intensified their search for NEOs, IfA's David Tholen upbraided colleagues for turning a blind eye to asteroids lurking inside Earth's orbit. He was concerned that an inner-orbit NEO at its farthest point from the sun could hit our planet. "For years, I wanted to do something about that," says Tholen. But he lacked the means. "Other folks had great cameras. I was envious." In 1997, he finally got time on a decent telescope. Aiming it

low on the horizon just after nightfall or before dawn, his group over 3 years discovered four asteroids in this blind spot—including a whopper that is 5 kilometers wide.

Riding high, Tholen won a grant for a more intensive search campaign. But his team struggled with technical glitches, and by their final year of funding in 2004, he says, "we hadn't found a single asteroid." He redoubled his efforts, booking time at observatories around the world. In June 2004, he was juggling nights on two telescopes. Then in the early evening of the 18th, at Kitt Peak National Observatory near Tucson, Arizona, Tholen, Roy Tucker, and Fabrizio Bernardi hit pay dirt: They got a first glimpse of Apophis.

"Apophis demonstrated that we know very little about the region of space near Earth," says Boris Shustov, director of the Institute of Astronomy in Moscow. Anxiety will mount when Apophis chugs back into range in 2012. Ironically, the best instrument for refining the asteroid's orbit—the world's most powerful planetary radar at Arecibo Observatory in Puerto Rico—may be switched off in 2011, the victim of budget cuts. Even without Arecibo, optical measurements almost certainly will reduce or rule out the impact risk. For that reason, NASA has no plans to send a probe to Apophis, and the European Space Agency has shelved a mission (see sidebar, p. 1329).

A chilling reassessment of Apophis could change the political landscape fast.

Suppose that observations forecast a 1-in-1000 impact risk in 2036. "That risk is really low, but if it hits, it's really bad," says Lu. "How much is it worth to us to have peace of mind?"

The "threshold of pain," as Lu calls it, may depend on who would be affected—and what resources they have. Based on current calculations, the line where Apophis might hit—the so-called risk corridor—runs from Kazakhstan through Siberia, over the northern Pacific, and across Costa Rica, Colombia, Venezuela, and the south Atlantic. Who would mount and pay for a deflection mission? All countries along the corridor? Just Russia, vulnerable to a direct hit, or the United States, vulnerable to a towering tsunami? The United Nations? What if a mission failed, deflecting Apophis to another point on the risk corridor, converting an "act of God" into an act of humankind? Who would be liable?

As experts grapple with these questions, some are trying to rouse political leaders. With outside advice, the Association of Space Explorers, an organization of astronauts and cosmonauts based in Houston, Texas, is drafting an NEO Deflection Decision Protocol to present to the U.N.'s Committee on the Peaceful Uses of Outer Space in 2009. "Apophis

## THE STATE OF OUR PLANET'S DEFENSES

Experts can't say exactly when the next Earth-bound asteroid will heave into view, but they are confident that humanity has the tools to defend itself. There are several deflection scenarios; in most, the straightforward objective would be to change an asteroid's speed so that it arrives too early or too late to hit Earth.

In 2005, former astronaut Edward Lu of Google and astronomer-astronaut Stanley Love, a mission specialist on the shuttle flight last month that delivered the Columbus Laboratory to the space station, proposed a "gravity tractor": a spacecraft that hovers in front of or behind an asteroid, using its minuscule gravitational force to slightly accelerate it or slow it. The gravity tractor could divert an asteroid from a keyhole, a narrow swath of space where gravitational forces would yank an asteroid onto a trajectory in which it would hit Earth a few years later.

Steering an asteroid clear of a keyhole would require less energy and thus is much easier to accomplish than diverting an asteroid on a direct course for Earth. A 1-ton gravity tractor would have to hover more than 3 years near a Tunguska-size, 45-meter-wide NEO on a collision course to change its orbit enough to bypass Earth, says Russell Schweickart, chair of the B612 Foundation. It would take less than 40 days to divert the much-larger Apophis from a keyhole on its close encounter with Earth in 2029, he says.

Another way to fiddle with Apophis's speed would be to spray it with material that changes the amount of sunlight it absorbs or reflects. If such a mission were mounted by 2018, just a few-percent change in its energy balance over 18 years would assure that Apophis misses Earth, says Jonathan Giorgini, a senior analyst at NASA.

Coaxing an asteroid to miss a keyhole won't necessarily eliminate the risk. Space is littered with keyholes and resonant return points that can sling an object back at Earth. To counter that possibility, scientists are devising ways to slap a transponder on a target asteroid or otherwise send back telemetry revealing whether a follow-up mission is necessary.

Last week, The Planetary Society in Pasadena, California, announced the winners of a competition to design a mission that would tag a potentially hazardous near-Earth asteroid to better track its orbit. The top prize of \$25,000 went to two companies—SpaceWorks Engineering Inc. in Atlanta, Georgia, and SpaceDev Inc. in Poway, California—for Foresight, a \$137 million spacecraft that would shadow Apophis for almost a year, taking its measure with a multispectral imager and a laser altimeter. The European Space Agency (ESA) has a similar mission on the drawing board: Don Quijote, which for €150 million would send a spacecraft to rendezvous with Apophis and precisely measure its position, mass, and other parameters. But Don Quijote is on hold. "Without an imminent impact

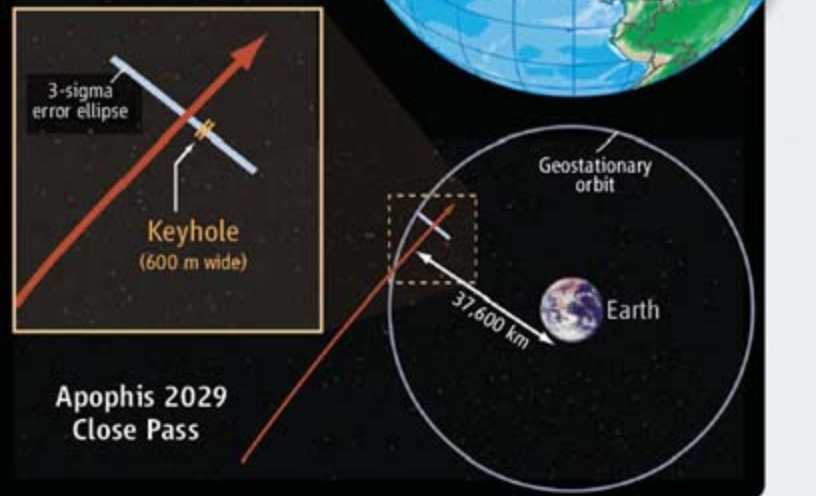
should unite our efforts to deal with the threat," says Shustov, who is leading an effort to develop Russia's first national R&D program on NEO hazards.

Shustov's nightmare is that leaders will drag their feet until the threat of a direct hit becomes real. But an asteroid need not impact to cause chaos. Each year, military satellites detect several 1-kiloton explosions of asteroids in the upper atmosphere, and every several years, a much larger explosion of 10 kilotons or more, says Sandia's Boslough. "They are quite frightening to people on the ground." A bus-size meteoroid would explode in the stratosphere with the energy of a small atomic bomb, producing a blinding flash much brighter than the sun, says Chapman. "Mili-

tary commanders in a region of tension might regard it as the hostile act of an enemy and retaliate," he says. A 25-kiloton airburst occurred over the Mediterranean Sea on 6 June 2002. Imagine, Chapman says, "if that had happened instead in the vicinity of Kashmir, where tensions between India and Pakistan were elevated."

While this scenario may argue for giving NEO sightings wide publicity, some experts think that detailed predictions—particularly risk corridors—should be withheld from the public. They want to avoid a "Chicken Little" phenomenon of repeatedly sounding alarms that are later downgraded or called off. NASA has not released Apophis's risk corridor in 2036. (The B612 Foundation provided the diagram

**Threading the needle.** Based on what is now known about Apophis's orbit, there is a slight chance that the asteroid will pass through a keyhole in 2029 that would put it on course for a collision in 2036, somewhere along the risk corridor at right.



threat, ESA is focusing on other priorities," says the agency's Ian Carnelli.

A modest-sized asteroid spotted too late for subtle maneuvering could be rammed with a kinetic impactor—a missile, or a spacecraft similar to NASA's Deep Impact probe that crashed into comet 9P/Tempel in 2005. According to a report last year in the journal *Science and Global Security*, a single strike, with 5 years' lead time, should safely divert an asteroid up to 250 meters wide; many larger asteroids could be deflected with multiple strikes and longer lead times. A gravity tractor in place before the strike could both provide telemetry and give the asteroid an extra nudge, if necessary.

Some scientists, meanwhile, favor a nuclear detonation, the force of which could obliterate smaller asteroids—with the hope that any fragments still on target for Earth would burn up in the atmosphere—or alter the trajectory of larger ones. Schweickart, for one, views nukes as a last resort. An asteroid for which other technologies would fail, he contends, comes along only once every 100,000 years or so. However, notes Harvard University astronomer Brian Marsden, "if the warning time is such that an object will hit us in a matter of months or even weeks, the nuclear option is the only one we really have. If the warning time is only days, I really don't know what we would do." —R.S.

above.) "We do not generally release these kinds of diagrams when they relate to future and ongoing risk assessments," says Steven Chesley, an NEO specialist at NASA's Jet Propulsion Laboratory in Pasadena, California.

Others believe in full disclosure. "People don't like secrecy. It breeds distrust," says Chapman. "When the facts are finally revealed, people wonder whether to believe them and wonder about what else might be still under wraps." NEO impact forecasts, he says, should be treated like hurricane forecasts, allowing people to respond.

Like the first hurricane of the season, the first test of our planetary defenses may be an asteroid whose name starts with the letter "A."

—RICHARD STONE





◀ **Victims of a hit?** Published evidence that an impact triggered the mammoths' disappearance is falling far short of proof.

### Diamonds not forever

Everyone agrees on one point at least. "Obviously, something really interesting happened 13,000 years ago," as Kring puts it. It was 12,900 years ago, to be precise, that a world staggering out of the last Ice Age suddenly plunged back into a millennium of near-glacial climate before emerging into the current warmth. It was also about then—emphasis on the uncertainties summed up by "about"—that the mammoths and other great beasts disappeared from North America. And the Paleo-Indian Clovis culture vanished from the archaeological record around then, too.

The *PNAS* authors have a cosmic explanation for the coincidence of climate shift, extinctions, and cultural oblivion: A body or clump of bodies from outer space ravaged North America. By exploding over or actually hitting the great ice sheet in the north, their reasoning goes, the impactors could have shifted climate into the chill of the so-called Younger Dryas (YD) period. And the blast or blasts, as well as the resulting continent-wide wildfire, would have sufficed to wipe out or at least seriously weaken man and beast.

Headed by nuclear chemist Richard Firestone of Lawrence Berkeley National Laboratory in California and retired geophysical consultant Allen West of Dewey, Arizona, the 26 *PNAS* co-authors present what they argue is debris from the impact: metallic bits, an abundance of the exotic element iridium, nanodiamonds, and molecular "buckyballs" filled with extraterrestrial helium. And the wildfire would have left charcoal, soot, carbon spherules, and glasslike carbon. Along with the impact debris, these components appear in a thin layer of sediments—the YD boundary layer—that was laid down near the beginning of the cold snap and the end of the mammoths.

That sort of litany impressed the largely nonexpert crowd at last May's Joint Assembly of the American Geophysical Union (AGU) in Acapulco, Mexico, but the few experts there were nonplussed. Now, in the wake of the detailed *PNAS* paper, the experts are able to take a more critical look. For starters, they are pointing out that the carbon-rich debris says nothing about the cause of the fires. Fire happened back then, notes geologist Nicholas Pinter of Southern Illinois University (SIU) in Carbondale, especially once humans arrived. Critics are equally quick to set aside the helium-filled buckyballs or fullerenes reported in the *PNAS* paper by geochemist Luann Becker of the University of California,

## PALEONTOLOGY

# Experts Find No Evidence for a Mammoth-Killer Impact

A devastating cosmic collision 13,000 years ago continues to play well in the media, but specialists are challenging the grounds for thinking it happened

It looked impressive as slide after data-laden slide flashed on the screen last spring. Nearly a dozen debris markers, found at 26 sites from the U.S. West Coast to Belgium, testified to a huge impact followed by a continent-spanning wildfire. The catastrophe had taken place a geologic instant ago—closely coinciding with the disappearance of North America's mammoths and the continent's earliest human culture (*Science*, 1 June 2007, p. 1264). Then came the 26-author paper last October in the *Proceedings of the National Academy of Sciences (PNAS)*, not to mention the hourlong National Geographic Channel documentary running on cable since last October, with more coverage on the way from the History Channel and PBS's prestigious program *NOVA*.

Although cosmically blasted mammoths may make good copy, many impact specialists have lately swung from leeriness to thorough disbelief. "The whole thing is contrived," says geochemist and impact specialist Christian Koeberl of the University of Vienna, Austria. "Their data don't agree with anything we know about impacts. It just doesn't make any sense. Occam's razor has been put safely in a drawer somewhere."

One problem is that no one has "any of the classic evidence of an impact," says impact specialist David Kring of the Lunar and Planetary Institute in Houston, Texas. Spurred by the 1980s debate over what killed off the

dinosaurs, "the community learned a lot about what the threshold of evidence is" for confirming an impact, he explains. But taking all the evidence offered by the group proposing the mammoth-killer impact, "you end up with [markers] that are not diagnostic of impact," says impact specialist Bevan French of the National Museum of Natural History in Washington, D.C. Proponents, meanwhile, are defending some of their published claims and giving ground on others but promising ultimate vindication.



**ET?** An impactor (top) may have produced magnetic spherules (lower right), but similar spherules (lower left) continually fall from space.

CREDITS (TOP TO BOTTOM): GIANNI DAGLI ORTICORRIS; STOCKTREK IMAGES/GETTY IMAGES; L. FRANZEN, N. PINTER, AND S. ISHMAN, GSA TODAY (2008); FIRESTONE ET AL., PNAS 104, 41 (2007)

Santa Barbara (UCSB). Throughout a half-dozen years of effort, no one else has replicated the isolation of fullerenes with helium (*Science*, 14 May 2004, p. 941).

Then there are the nanodiamonds. Zillions of diamond bits a few nanometers in size sound exotic enough. Many meteorites are filthy with them, so the impactor could have brought them in. Nanodiamonds have in fact been reported in the debris of the dinosaur-killing impact 65 million years ago.

At the AGU meeting, paleoceanographer and *PNAS* third author James Kennett of UCSB reported that UCSB colleagues had “conclusively” shown the presence of nanodiamonds in sediments from the YD boundary layer. They used transmission electron microscopy (TEM), the gold standard for nanodiamond identification. However, no TEM results appeared in the *PNAS* paper. Instead, a sample of glassy carbon recovered from the YD boundary had been sent to a commercial laboratory for analysis using carbon-13 nuclear magnetic resonance (NMR). The NMR analysis showed that the “sample contains nanodiamonds, which are inferred to be impact-related material,” the paper states.

Experts asked to comment on the findings disagree. “Their NMR data do not provide evidence for nanodiamonds,” says geochemist George Cody of the Carnegie Institution of Washington’s Geophysical Laboratory in Washington, D.C., who in 2002 was the first to use NMR to identify nanodiamonds in meteorites. “I would never have claimed that [their NMR spectrum] had anything to do with nanodiamonds.”

Under the proper analytical conditions, says Cody, nanodiamonds produce a narrow NMR peak centered at a chemical shift of 34 parts per million. The *PNAS* spectrum is broad and centered at 38 parts per million, too broad and too far afield to be nanodiamonds, he says. In any case, the analytical conditions used were wrong for detecting nanodiamonds, Cody adds; no peak would have appeared even if they were there.

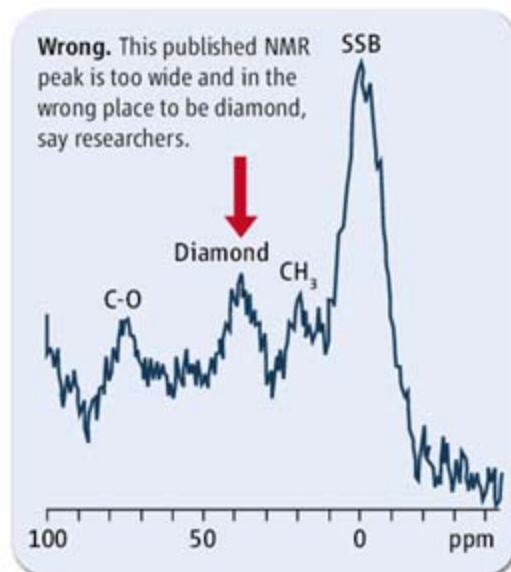
### Mundane metals?

Another claimed marker of the YD impact—the element iridium—is coming under attack as well. An iridium “spike” was the first clue to identifying the impact that caused the Cretaceous-Tertiary (K-T) mass extinction 65 million years ago. The metallic element is scarce in Earth’s crust but relatively abundant in meteorites, so like nanodiamonds, any excess might have arrived via asteroid or comet.

Firestone and colleagues reported elevated iridium of a few parts per billion (ppb)—

comparable to K-T sediments—in some sediment samples from the YD boundary but not in sediments above or below. They found tens to more than 100 ppb of iridium in microscopic particles—both rough grains and once-melted spherules—magnetically separated from some of those sediments. And they cite an earlier report in *Nature* of “large increases” in iridium “during the Younger Dryas as recorded in the GRIP (Greenland) ice core.” The iridium came from beyond Earth in an impactor, the group concluded.

Other researchers aren’t sure where the iridium came from, if it’s there at all. As to the ice core record, “I was surprised to see such an interpretation of our results in *Nature*,” says Paolo Gabrielli, first author of the *Nature* paper and now at Ohio State University in Columbus. “My paper does not report any large increase of iridium in the Younger Dryas.



So it has nothing to do with an extraterrestrial impact,” Firestone disagrees: “I interpret his results differently than he does.”

Impact specialist Philippe Claeys of the Free University of Brussels in Belgium can’t find any iridium at all in the four sediment samples of the YD boundary West sent him for analysis. The *PNAS* group eventually reported that two of the samples contained elevated iridium easily detectable by Claeys’s method; the magnetic fraction of the third sample had extreme iridium concentrations. But Claeys reported to West that he could detect no iridium higher than 0.5 ppb in any of the samples. West blames the “nugget effect,” in which a few microscopic sediment particles highly enriched in iridium account for most of the iridium in an analyzed sample; samples that happen to have few nuggets look barren. Claeys, however, says he intentionally used large enough samples to avoid the nugget effect.

Archaeologist Vance Haynes, professor emeritus at the University of Arizona, Tucson, is finding likely looking magnetic spherules in the darnedest places. He has spent 30 years studying Clovis sites, many of which the Firestone group sampled. As a check on his own ongoing independent analysis of YD samples, he collected a modern sample. “I got 300 grams of dust off the roof [of my house], and it’s full of magnetic microspherules,” he says. Whether they are the melted, iridium-rich micrometeorites that continually drift down from the upper atmosphere or the product of high-temperature industrial processes such as coal burning, he doesn’t yet know. Either way, they could be trouble. The cosmic dandruff of microspherules could have salted sediments forming 12,900 years ago with iridium, while the humanmade variety might have settled on modern outcrops before sampling.

Chemical analyses of the magnetic particles do not point to impact, Koeberl says. The elemental analyses make little geochemical sense, he says. In particular, the magnetic particles are far too rich in titanium to be extraterrestrial. He rejects the suggestion in the *PNAS* paper that such odd geochemistry points to “a new and unknown type of impactor.” Meteoriticist Theodore Bunch of Northern Arizona University in Flagstaff, the fifth *PNAS* author, agrees that the magnetic fraction has problems. What its chemistry means, “I don’t know,” he says, speaking for himself. In any case, “it detracts from the main thing.”

The main thing now is nanodiamonds, according to Bunch and other *PNAS* authors. The initial UCSB detection of nanodiamonds came too late for their paper, says Firestone. Now West is using TEM and has found three different types of nanodiamonds in the YD layer but failed to find any above or below it. “Some people just can’t stand the idea of something falling out of the sky,” he says, but “they can’t explain all of these [impact] markers, and diamond is the hardest to explain away.”

West and colleagues expect to publish on nanodiamonds, but their critics are still waiting to be impressed. Pinter and Scott Ishman, his micropaleontologist colleague at SIU, wrote in a detailed critique in the January issue of *GSA Today* that such “spectacular stories to explain unspectacular evidence consume the finite commodity of scientific credibility.” The problem, Pinter says, is that “there’s a wide fringe beyond the impact community” where the criteria for impact identification laid out in the literature are not rigorously followed. Whether another try at nanodiamonds will meet the standard is anybody’s guess.

—RICHARD A. KERR



## PLANT SCIENCES

## Corn Genomics Pops Wide Open

The sequencing of maize genomes and the development of new strains are enabling faster exploitation of this key crop's natural diversity

A decade ago, sequencing the maize genome was just too daunting. With 2.5 billion DNA bases, it rivaled the human genome in size and contained many repetitive regions that confounded the assembly of a final sequence. But last week, not one but three corn genomes, in various stages of completion, were introduced to the maize genetics community. In addition, researchers announced the availability of specially bred strains that will greatly speed tracking down genes involved in traits such as flowering time and disease resistance. These resources are ushering in a new era in maize genetics and should lead to tougher breeds, better yields, and biofuel alternatives. "We're sitting on very exciting times," says Geoff Graham, a plant breeder at Pioneer Hi-Bred International Inc.

The world's biggest crop, maize (*Zea mays*) comes in all shapes and sizes. Indeed, the genomes of any two varieties can be as different as chimp and human DNA. Cataloging, understanding, and harnessing this variation to improve crop yields have been longtime goals for researchers.

Toward that end, in 2005, the U.S. National Science Foundation (NSF) and the U.S. departments of Agriculture (USDA) and Energy (DOE) provided \$30 million to a consortium headed by Richard Wilson at Washington University in St. Louis, Missouri, to tackle the genome of a well-studied maize strain called B73. Rod Wing of the University of Arizona, Tucson, provided 15,000 mapped segments of the corn's DNA for sequencing, and at a meeting\*

\* "The 50th Annual Maize Genetics Conference," 27 February–2 March, Washington, D.C.

last week in Washington, D.C., Wilson described B73's draft genome. About 6500 of the segments Wing provided are completely finished and most of the rest are well under way. Even at this stage, "we believe the quality and coverage will enable new discoveries," says Wilson.

Maize researchers agree. B73's full sequence "is going to underpin all the research that we do in maize genomics," predicts Patrick Schnable of Iowa State University in Ames.

Take the quest to improve the potential of corn and perennial grasses as biomass for biofuels. A key goal is to increase sugar content and sugar's availability for conversion to biofuels. "We need to greatly increase mass per acre," says Nicholas Carpita, a plant cell biologist at Purdue University in West Lafayette, Indiana. He and his colleagues have compared the rice and *Arabidopsis* genomes with the B73 DNA already deposited in the public database GenBank. They found more than 1400 corn genes involved in building plant cell walls—the ultimate energy sources—and are homing in on those that affect biomass quantity and quality. "The maize genome allowed us to get to [those] genes," he says.

And the B73 genome isn't the only one in the works. With \$9.1 million from the Mexican government, Jean-Philippe Vielle-Calzada of the National Laboratory of Genomics for Biodiversity in Irapuato and his colleagues have decoded a native "popcorn" strain grown at elevations above 2000 meters. Although still in more than 100,000 pieces, the sequence has revealed many new genes, he reported. This variety's genome "will be of tremendous value

**Field tech.** Bar-coding tools speed maize genetics research.

in terms of understanding the evolution of [maize] domestication," he says.

In addition, Daniel Rokhsar of DOE's Joint Genome Institute in Walnut Creek, California, and his colleagues have begun to decipher the DNA of a well-studied maize strain called Mo17, using new, cheaper sequencing technologies. If the effort proves cost-effective, NSF and DOE may support the sequencing of additional strains.

But genome sequences aren't the only big news for maize researchers. As part of the Maize Diversity Project, USDA plant geneticist Edward Buckler of Cornell University and his colleagues have bred almost 5000 lines of maize, revealing the full range of the plant's diversity. These lines were derived from crosses between B73 and 25 other inbred maize lines from all over the world; each marriage has given rise to about 200 lines. For the past 2 years, teams have planted these lines in 11 fields across the United States and measured many different traits—height, cob size, flowering time, and so on—for each line.

Using those lines, Buckler's team has also put together a detailed genetic map of the maize genome, which is helping researchers home in on target genes by means of an approach called nested association mapping. "It's an incredible resource ... on equal par to having the sequence," says Cornell's Thomas Brutnell.

Using the map, researchers can easily pinpoint the spots on the genome that underlie variation in a particular trait, then use the genome sequence to figure out which gene is at that spot. "It holds [great] power," says Jay Hollick of the University of California, Berkeley. "Virtually any trait can be measured."

Already, Buckler reported, his team has pinned down 50 genes that dictate flowering time. Some lines flower as much as 45 days apart, but no single gene region shifted flowering time by more than 3 days.

Another resource introduced at the meeting will help Buckler and others sort out how genes interact. The agribusiness giant Syngenta announced it was making available 7500 lines of corn, each representing a B73 genome with a single piece of DNA bred into it from one of the 25 strains of the Maize Diversity Project. Taken together, the lines incorporate all the genetic diversity of those strains but make it easier to understand the activity of particular genes. The community has long awaited these tools, says Brutnell: "They are really going to revolutionize the way we do genetics."

—ELIZABETH PENNISI

Fairly unbalanced

1338



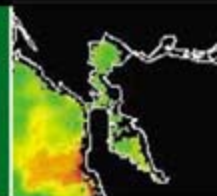
The age of the Grand Canyon

1343



Phytoplankton shifts

1346



LETTERS | BOOKS | POLICY FORUM | EDUCATION FORUM | PERSPECTIVES

## LETTERS

edited by Jennifer Sills

### Editorial Expression of Concern

IN THE 1 JULY 2005 ISSUE, *SCIENCE* PUBLISHED THE REPORT "A MAGNETIC NANOPROBE TECHNOLOGY FOR DETECTING MOLECULAR INTERACTIONS IN LIVE CELLS" BY J. Won *et al.* (1). Professor Gyun Min Lee, Chair, The Internal Investigation Committee, Department of Biological Science, Korea Advanced Institute of Science and Technology (KAIST), notified *Science* on 28 February 2008 that this article, published by Professor Tae-Kook Kim and his co-workers, is being investigated along with a paper published in *Nature Chemical Biology* (2). The correspondence from Professor Lee states that, although the formal investigation has not yet been completed, "our initial investigative results are strong enough to convince us that the two papers do not contain any scientific truth."

*Science* is publishing this Editorial Expression of Concern to alert our readers to the fact that serious questions have been raised about the validity of the findings in the Won *et al.* paper. We are working with the authors and KAIST to determine appropriate next steps.

DONALD KENNEDY<sup>1</sup> AND BRUCE ALBERTS<sup>2</sup>

<sup>1</sup>Editor Emeritus; <sup>2</sup>Editor-in-Chief

#### References

1. J. Won *et al.*, *Science* **309**, 121 (2005).
2. J. Won *et al.*, *Nat. Chem. Biol.* **2**, 369 (2006).

### Rewarding Reviewers

DOUBTLESS, HORDES OF EDITORS NODDED their heads in commiseration with W. F. Perrin's Letter, "In search of peer reviewers" (4 January, p. 32), weary of both difficulty in recruiting competent peer reviewers and complaints from the same reluctant professionals about the inadequacy of their own manuscripts' reviewers. In my experiences editing a book and managing numerous proposal review panels for two government agencies, Perrin's figure of "8 or 10 tries to find someone" rings true. But his admonishment of scientists who are evasive to review yet quick to complain lacks a measure of introspection that could drive real change.

In the crucible of academic advancement, scientists have staggering demands on their time. Perrin's suggestion to include reviewing in job descriptions and tenure evaluations acknowledges that these pressures divert scientists from volunteering reviews. However, editorial practices generally lack the recognition needed to make serving as a reviewer a

plausible metric of performance, and provide little positive feedback for the effort. In my experience with *Science*, submitting a thoughtfully crafted review elicited no acknowledgment. Was the review received? Were the comments instructive? What was the fate of the manuscript in question?



Editors will best succeed in getting reviewers not by simply making reviewing a requirement, but by doing their part to see that it is rewarded. Even an acknowledgment adding the reviewer's name and journal's manuscript number to a form letter would make the activity "count." A somewhat more

gracious reply would assuredly appeal to the desire prevalent among scientists to serve the community.

The editorial establishment could go even further to encourage willing reviewers at little added expense. They could leverage a primary currency of academic science—prestige—and present an award to their best reviewer(s) each year. They could also help make reviewing a component of

researchers' competitiveness for funding by encouraging funding agencies to include a count of average manuscript reviews per year on applicant CVs.

MATTHEW A. METZ

CUBRC Center for International Science and Technology Advancement, Washington, DC 20036, USA.

### Rating Reviewers

AS THE EDITOR-IN-CHIEF OF *ACM TRANSACTIONS ON INFORMATION SYSTEMS* and a member of a dozen editorial boards over the years, I resonate with the Letters by W. F. Perrin and R. S. Zucker (4 January, p. 32), and I encourage even more vigorous consideration of peer reviewer responsibilities and merits. Good reviews are a precious human resource for the advancement of science, and we are beginning to see more accountability and consideration of reviews on several fronts. Traditional peer-reviewed journals are increasingly adopting electronic manuscript management systems, which provide databases of reviews that persist beyond the tenures of individual editors. This technology offers new challenges and opportunities to assess the review process. Not only can reviews be stored, managed, and possibly aggregated, but many of the commercial and professional journal systems allow editors to rate the quality of reviews (although at present, this typically consists of simple Likert-scaled values for timeliness and depth). These systems will increasingly lead to tangible ways to at least count reviewer participation and perhaps assess the quality of reviewer participation in a scientific community over time. Editors may be asked to provide assessments of review quality for promotion and tenure committees or immigration appeals. Such systems also have the potential to make ephemeral blind reviews more archival, thus providing a historical record of the reviews themselves and a possible corpus of data for future historians of science to mine in order to understand how reviews fit into the development of scientific breakthroughs and human progress. Questions about review ownership may also arise.

Reviewers should be encouraged to actively participate in the scholarly discourse of publication and be rewarded for this participation. Young researchers must especially

understand that their participation is not only expected, but that the ability to assess this participation will increase over time.

GARY MARCHIONINI

School of Information and Library Science, University of North Carolina, Chapel Hill, NC 27599, USA.

## Preventing Inequity in International Research

BEING ORIGINALLY FROM AN EASTERN European country, I've noticed two possible practices in establishing international research collaboration between richer and poorer countries. In the more desirable scenario, investigators from wealthy countries spend time living in poorer countries, where they patiently gain trust of local people and build capacity in local research infrastructure. In the other scenario, investigators use local researchers to perform the difficult, risky, and demanding part of the work, after which they collect the raw data and begin to publish papers. In my international health work in Africa, I learned of the extremes of both approaches. The latter breed of scientists was nicknamed the "vampires" by the local community, as they were only seen when flying in to collect blood samples. It was difficult to avoid associations to the "vampires" when reading the Report "Cognitive recovery in socially deprived young children: The Bucharest Early Intervention Project," by C. A. Nelson III *et al.* (21 December 2007, p. 1937).

Reading this Report raised a number of questions. How was it possible to publish this study without any coauthors based at Romanian institutions? If this study is not cause for major ethical concerns, as a related Policy Forum (1) suggested, why was it not initially conducted in deprived areas of the western country? Would *Science* publish exactly the same paper on 187 Romanian children if all six coauthors were Romanian scientists?

I am absolutely sure that the authors of the study and the *Science* editors could easily provide perfectly reasonable answers to all three of my intuitive questions. However, I don't

### Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 3 months or issues of general interest. They can be submitted through the Web ([www.submit2science.org](http://www.submit2science.org)) or by regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before publication. Whether published in full or in part, letters are subject to editing for clarity and space.

### TECHNICAL COMMENT ABSTRACTS

#### Comment on "Early Archaean Microorganisms Preferred Elemental Sulfur, Not Sulfate"

Huiming Bao, Tao Sun, Issaku Kohl, Yongbo Peng

Philippot *et al.* (Reports, 14 September 2007, p. 1534) interpreted multiple-sulfur isotopic compositions of ~3.5-billion-year-old marine sulfide deposits as evidence that early Archaean microorganisms were not sulfate reducers but instead metabolized elemental sulfur. However, their data can be better explained by a scenario involving poor mixing of photochemical and surface sulfide sources.

Full text at [www.sciencemag.org/cgi/content/full/319/5868/1336b](http://www.sciencemag.org/cgi/content/full/319/5868/1336b)

#### Response to Comment on "Early Archaean Microorganisms Preferred Elemental Sulfur, Not Sulfate"

Pascal Philippot, Mark Van Zuilen, Kevin Lepot, Christophe Thomazo, James Farquhar, Martin J. Van Kranendonk

Our knowledge of the sulfur cycle on early Earth is still in its infancy. Nevertheless, there exist enough geochemical constraints from the rock record to show that the theoretical mixing models proposed by Bao *et al.* are highly unlikely to account for the range of  $\delta^{34}\text{S}$  and  $\Delta^{33}\text{S}$  values recorded for the microscopic sulfides at the North Pole.

Full text at [www.sciencemag.org/cgi/content/full/319/5868/1336c](http://www.sciencemag.org/cgi/content/full/319/5868/1336c)

### CORRECTIONS AND CLARIFICATIONS

**News of the Week:** "A plan to capture human diversity in 1000 genomes" by J. Kaiser (25 January, p. 395). Wang Jun is associate director of the Beijing Genomics Institute, Shenzhen.

**Research Articles:** "Correlation of the highest-energy cosmic rays with nearby extragalactic objects" by The Pierre Auger Collaboration (9 November 2007, p. 938). An author name was omitted from The Pierre Auger Collaboration: B. Kégl is at the Laboratoire de l'Accélérateur Linéaire, Université Paris-Sud, IN2P3/CNRS, Orsay, France.

think that is really the point. The point is that when international research collaborations between the scientists from the wealthier and the poorer countries are based on good principles of equity and mutual respect, questions like these should never even come to mind.

IGOR RUDAN

Croatian Centre for Global Health, University of Split Medical School, Split, Croatia.

#### Reference

1. J. Millum, E. J. Emanuel, *Science* **318**, 1874 (2007).

#### Response

WE AGREE WHOLEHEARTEDLY WITH RUDAN that U.S.-based investigators should invest themselves in the local community. As noted in the Supporting Online Material, we have established an Institute of Child Development in Bucharest. The Institute is designed to work collaboratively with governmental and non-governmental agencies to improve the lives of Romanian children and train the next generation of Romanian psychological and biomedical researchers. Similarly, the three main investigators of this project (C.H.Z., N.A.F., and C.A.N.) have appointments on the faculty of the Department of Psychology at the University of Bucharest, where they lecture several times a year and supervise masters and doctoral students. Additionally, we raised the funds to support two large national conferences, each of which brought more than 300

professionals from all over Romania together to learn from international experts about child development, neuroscience, and child protection. Finally, we have invested in the continuing education of our staff and colleagues in Romania by financially supporting their travel inside and outside of Romania to attend conferences and workshops.

Rudan asks why this work was not done in the United States or a similar "western" country. We address this in some detail in our Supporting Online Material. Briefly, there are very few nonhandicapped institutionalized children in such countries, whereas in Romania, there were tens of thousands of children being raised in institutions. In addition, we conducted the study in Romania because the Secretary of State for Child Protection originally invited us to conduct the study.

Rudan asks why none of our Romanian colleagues were authors of this paper. As we note in reference 32, we are indebted to many individuals in Romania who facilitated our research. We began the study with collaborators who were primarily involved in child protection or clinical work—not child development research. These colleagues did not make the requisite substantial scholarly contributions to this paper required by scientific organizations (such as the American Psychological Association) for authorship. That is, none contributed to the experimental design, statistical

analysis, data interpretation, or manuscript preparation that one typically associates with journal authorship. This does not diminish their contributions to the study in any way, and indeed, three of our most important colleagues are authors on another paper from the study (1).

Rudan advocates that the investigators spend a substantial amount of time in poorer countries before conducting research. He contrasts this with exploitation of local investigators by outsiders. We submit that there may be other alternatives. As desirable as Rudan's first approach may be, our resources did not permit us to adopt it. Instead, we hired an all-Romanian staff, invested in their training, and had a full-time Romanian-American project manager who lived in Bucharest for the project's first 4 years. In the past 7 years, we have continued to seek and nurture collaborations with academics and child protection professionals, and we hope that those will continue to grow over time. In addition to the Institute, the legacy of BEIP includes creation of the largest foster care network in Romania at the time, supported fully by project funds at the outset, but later transferred to the gov-

ernment. These foster homes continue to provide care to abandoned children. Thankfully, they have now been joined by hundreds of new foster homes throughout Bucharest and beyond.

**CHARLES A. NELSON III,<sup>1\*</sup> CHARLES H. ZEANAH,<sup>2</sup>  
NATHAN A. FOX,<sup>3</sup> PETER J. MARSHALL,<sup>4</sup>  
ANNA T. SMYKE,<sup>2</sup> DONALD GUTHRIE<sup>5</sup>**

<sup>1</sup>DMC Lab of Cognitive Neuroscience, Harvard Medical School and Children's Hospital, Boston, MA 02215, USA. <sup>2</sup>Tulane University Health Sciences Center, New Orleans, LA 70112, USA. <sup>3</sup>Department of Human Development, University of Maryland, College Park, MD 20742, USA. <sup>4</sup>Department of Psychology, Temple University, Philadelphia, PA 19122, USA. <sup>5</sup>University of California, Los Angeles, CA 90095, USA.

\*To whom correspondence should be addressed. E-mail: charles.nelson@childrens.harvard.edu

#### Reference

1. C. H. Zeanah et al., *Infant Mental Health J.* **27**, 559 (2006).

## AIDS Vaccine Research: Consider Co-Infections

IN THE NEWS OF THE WEEK STORY "TRIALS OF NIH'S AIDS vaccine get a yellow light" (21 December 2007, p. 1852), J. Cohen and B. Lester report on the NIH Vaccine

Research Center (VRC) AIDS vaccine trial that was put on hold, and they discuss whether the initiative should be continued. The VRC AIDS vaccine is based on an adenovirus 5 (Ad5) vector similar to the Merck STEP vaccine for which the trial was halted earlier this year (1). The authors cite the arguments for and against continuing the VRC trial in a smaller group of participants that excludes people who have antibodies against Ad5. The arguments include ethical issues inherent in a vaccine with limited applicability and the lack of proven efficacy of the vaccine in preclinical research.

We suggest another factor to consider in making this decision: The risk for the Ad5 seronegative participants to become infected with both HIV and Ad5 should be taken into account, especially in countries, such as Swaziland, where both viruses are highly endemic (seroprevalence 25% and 90%, respectively).

**ROB GRUTERS AND AB OSTERHAUS**

Department of Virology, Erasmus MC, s'Gravendijkwal 230, Rotterdam, Netherlands.

#### Reference

1. J. Cohen, B. Lester, *Science* **318**, 1852 (2007).

## What makes a first-class news story?



**Constance Holden**  
2004 National Mental Health Association: Media Award



**Richard Kerr**  
2006 Geological Society of America: Public Service Award



**Colin Norman**  
News Editor, Science magazine



**Jennifer Couzin**  
Articles selected for inclusion in The Best American Science Writing 2007 and 2005  
2003 Evert Clark/Seth Payne Award for Young Science Journalists

**A first-class editorial team.** Award-winning journalists write for *Science*—with 12 top awards in the last four years. That's why we have the most compelling stories, and the biggest readership of any general scientific publication. To see the complete list of awards go to:

[sciencemag.org/newsawards](http://sciencemag.org/newsawards)



## COMPUTERS

## News from Avatar Editors

Dan Hunter

Of the many ways to bifurcate the human population, one is to consider whether you will find the book *The Second Life Herald* compelling or silly. A simple way to decide this for yourself is to consider its opening paragraph:

It was a quiet night at the *Alphaville Herald*. The newspaper had been put to bed and a man known as Urizenus, its publisher, was as usual the last one in the office. He busied himself closing up shop, tending to fireplaces and cleaning up the messes that had accumulated over the course of the day. It took a few minutes to get Uri's cats, Cheddar Cheese Cheetah and a tabby named Black, back into their cages. Then he turned out the lights, locked up for the night, and headed home.

The events in that paragraph and much of what follows in the book took place in a virtual world. As any number of breathless news stories have by now explained, these worlds are online spaces where users navigate around a

imaginations of the authors and the other users of the computer systems that constitute the two worlds at the core of the book, *The Sims Online* and *Second Life*.

Apart from studying virtual worlds for a living, I am among the converted on this question, and I have no problem with the idea that these events are meaningfully real. Some genuinely important moments of my life have occurred within the boundaries of various online worlds. For people like me, who don't find the opening paragraph slightly ludicrous, the book is one of the best accounts of why these worlds matter now—and will matter even more in the future, as they are put to all manner of uses: scientific experiments, social interactions, work, property investment, sex, and so on. Peter Ludlow (a professor of philosophy and linguistics at the University of Toronto) and Mark Wallace (a freelance journalist)

**The Second Life Herald**  
The Virtual Tabloid That  
Witnessed the Dawn of  
the Metaverse

by Peter Ludlow and  
Mark Wallace

MIT Press, Cambridge, MA,  
2007. 311 pp. \$29.95, £19.95.  
ISBN 9780262122948.

form of the *Heralds'* editors or some aggrieved users whose cases they champion). And the book is written somewhat in the style of the *Heralds*, that is, a mannered impersonation of a hard-boiled detective/newspaperman circa 1930. For those familiar with *Second Life*, the authors' impersonation of J. J. Hunsecker (via Dashiell Hammett, as interpreted by Humphrey Bogart, with a dash of Cary Grant, etc.) will come as no surprise. *Second Life* is mostly an elaborate drag performance, and the acting out of various elaborations of the self—sexual, physical, emotional, and so on—is about the only thing

going for the environment. People sure as hell aren't there because of the aesthetic appeal of the environment.

[This doesn't quite explain why *Second Life* became, briefly, such a hotspot for corporate public relations. Nevertheless, it's evident that there was more going on there than the simple economics of hype and the fact that *Second Life* was a relatively cheap way for anxious marketers to head into online spaces that they didn't really understand. *Second Life*

## THE SECOND LIFE HERALD

Always Fairly Unbalanced.

representation of a physical space using their avatar or "toon." In doing so, they potentially interact with thousands or hundreds of thousands of other users. Although these environments are remarkable technological achievements, for all their sophistication they still demand a huge imaginative leap. They ask the user—and the authors of this book ask the reader—to "be" in the offices of this virtual newspaper. They ask us to take seriously the idea that those sets of scripted pixels that appear on the screen as "cats" aren't just entries in a database, but really are Urizenus's pets.

And so how you feel about this book will, in the end, come down to the metaphysical question of whether you think that the events recounted in it actually "took place" inside some world—rather than within the febrile

were the editors of the virtual newspapers the *Alphaville Herald* (in *The Sims Online*) and the *Second Life Herald* (in *Second Life*), and they take these worlds very seriously. The stories that they have collected from their eponymous newspapers—stories of virtual prostitution, asset trading, censorship, and the like—are amusing, often racy, and engaging. Their meditations on some of the interesting questions that emerge from virtual worlds—on issues like what experiences are real, or how we should structure the governance of these worlds—are always thoughtful and frequently insightful.

But for the many who don't buy the importance of virtual worlds, this book may just be the evidence they need to dismiss them altogether. The book's dominant theme is a meditation on the injustices that game developers inflict upon users. At times, the narrative comes across as an airless, quasi-Marxist diatribe, with big business (in the form of the developers) crushing the proletariat (in the



Urizenus Sklar, founder and contributing editor.

is all performative surface and doesn't reflect life in any serious sense (game worlds like *World of Warcraft* or *Everquest* are profoundly more mimetic). The thinness of the projections available in *Second Life*, and the plasticity of the images, must have appealed to public relations and marketing types. But that is a story for another day.]

In any event, readers unfamiliar with what Ludlow and Wallace are doing may find some tropes jarring. After a while, even I became irritated with the recounting of how Uri smoked a virtual stogie while raising a virtual glass of scotch to some competitor or how the *Herald* was so cheaply "printed" (on the Web)

The reviewer is at the University of Melbourne Law School, Australia, and the Legal Studies and Business Ethics Department, Wharton School, University of Pennsylvania, 3730 Walnut Street, Suite 600, Philadelphia, PA 19104-6340, USA. E-mail: hunterd@wharton.upenn.edu

CREDIT: COURTESY PETER LUDLOW/WWWW.SECONDLIFEHERALD.COM

that it left virtual ink stains on the hands of the residents of The Sims Online.

However, with luck and some imagination, readers will be able to look past these stylistic tics. They will find the book answers the question of why they should care about virtual worlds. There is life there, and it differs from life in physical spaces. Ludlow and Wallace offer what may be the best overview of the lives that are lived online. There are any number of books explaining how to make money in Second Life or how to “win” in the various game worlds, but until now there were only three seminal works—Edward Castronova’s *Synthetic Worlds (1)* and Julian Dibbell’s *My Tiny Life (2)* and *Play Money (3)*—that managed to help the general reader understand why these new worlds might matter. *The Second Life Herald* is a worthy addition to this small group and provides a useful, readable guide to the recent past and potential future of online worlds.

#### References

1. E. Castronova, *Synthetic Worlds: The Business and Culture of Online Games* (Univ. of Chicago Press, Chicago, 2005).
2. J. Dibbell, *My Tiny Life: Crime and Passion in a Virtual World* (Holt, New York, 1998).
3. J. Dibbell, *Play Money: Or, How I Quit My Day Job and Made Millions Trading Virtual Loot* (Basic Books, New York, 2006).

10.1126/science.1154771

## PUBLIC HEALTH

# Complexities of a Continent

Alash'le G. Abimiku

**H**IV/AIDS, *Illness, and African Well-Being* links history, cultural exchange, economic exploitation, and diseases across Africa in a very interesting and holistic manner that captivates the reader. The chapters were originally presented at a 2005 conference at the University of Texas at Austin, where editors Toyin Falola and Matthew Heaton are, respectively, professor and graduate student in the Department of History. The authors astutely place the havoc caused by HIV in Africa within the broader context of a continent struggling with infectious diseases, high rates of disability, and frequent conflicts

that substantially weaken its ability to achieve economic stability or respond appropriately to the HIV/AIDS pandemic.

From the very beginning of the book, one senses the interplay between historic events and dwindling public health: the decades of use of bactericidal “chewing sticks”; a missed opportunity in Ghana to educate the community on misconceptions about epilepsy; the compromised health of South African miners and the accompanying poverty, boredom, depression, alcoholism, and prostitution (all ingredients that further fuel infectiousness and hopelessness across the continent). A very fascinating link between social activities and health was forged by the globalization of hajj. Heaton’s chapter on the historical development of the West African pilgrimage scheme (1919–1938) clearly demonstrates how the global health system was compromised by a colonial regime more interested in staying in power than in protecting the health and citizenship of its pilgrims. However, colonization had some positive impact on Africans’ health (for example, the reduction of river blindness and eradication of smallpox).

By presenting Africa’s health issues in the context of its past socioeconomic practices, the book leads readers to envision better health outcomes that could have been based on the best of traditional and westernized Africa. One wonders, for example, how effective the eradication of malaria might have been if indigenous knowledge had been used along with Western medical discoveries. The overreliance on the latter led to overuse and underdosing, respectively, of DDT and chloroquine, which promoted resistant strains. Iruka Okeke’s especially informative chapter on the trends and containment of antimicrobial resistance in Africa offers potential solutions.

Quite appropriately, the editors devote several chapters to the devastation caused by HIV. These include quotes from key individuals such as Kofi Annan, Paul Farmer, Stephen Lewis, and Nobel laureate Joshua Lederberg, who point out that despite multilateral initiatives targeting the HIV pandemic in sub-



**Collaborative effort.** Johns Hopkins’s Laura Guay (left) and Makerere University’s Philippa Musoke at their lab building in Kampala, Uganda, the day that the country’s first shipment of nevirapine arrived.

### HIV/AIDS, Illness, and African Well-Being

Toyin Falola and Matthew M. Heaton, Eds.

University of Rochester Press, Rochester, NY, 2007. 428 pp. \$75. ISBN 9781580462402. Rochester Studies in African History and the Diaspora.

Saharan Africa the global response has been complacent. Yacouba Banhoro’s perspective on trends in AIDS in Burkina Faso between 1985 and 2001 questions the efficacy with which state policies and strategies to fight HIV/AIDS have been implemented. Mandi Chikombero reports focus-group discussions among Zim-

babweans that illustrate how a culture that does not allow open discussion of sex worsens the fear, anger, confusion, and denial experienced by HIV-infected persons. Such hopelessness could explain the deliberate attempts by infected persons to spread the virus to unsuspecting victims. Nevertheless, the successes enjoyed by Uganda and Senegal in reversing upward trends of AIDS clearly show that with a committed

government at the helm other African countries could at least contain their epidemics.

*HIV/AIDS, Illness, and African Well-Being* highlights the social, economic, and epidemiological complexities that beset the health of tens of millions of Africans. The authors’ accounts suggest that effective responses to epidemics such as HIV/AIDS in Africa must combine efforts to block disease transmission, improve public health, and promote economic development into a more holistic approach. Alliances between researchers in the developed West and their African colleagues to seek effective and lasting solutions to Africa’s diseases may offer lessons on how the West can care for neglected minority populations that share similar epidemics with Africa. Global efforts to improve Africans’ health must be bold, not tentative, and draw on lessons from history.

10.1126/science.1152217

The reviewer is at the Institute of Human Virology, University of Maryland, Division of Epidemiology and Prevention, 725 West Lombard Street, Baltimore, MD 21201, USA. E-mail: aabimiku@ihv.umaryland.edu



## MEDICINE

# Moving Toward Transparency of Clinical Trials

Deborah A. Zarin\* and Tony Tse

An almost steady flow of articles have focused on the dangers or lack of efficacy of widely used drugs, along with allegations of hidden information, misinterpreted data, regulatory missteps, and corporate malfeasance. Many of these accounts involve analyses of research on human volunteers that had never been publicly disseminated (1, 2). The uproar caused by an analysis of previously unpublished studies of the diabetes drug, Avandia, indicating that it may be harmful (3), is one recent example (4–8). As a result, many question whether sufficient information about the safety and efficacy of medical interventions is available to the public (9) and whether society is meeting its eth-

ical responsibilities to human volunteers who put themselves at risk (10, 11). Although advances in all areas of science depend on free exchange of data, clinical trials warrant particular scrutiny because of their use of human volunteers and our dependence on their results to inform medical decisions.

registration policies, which mandate the public listing of basic trial information, and results database policies, which mandate submission and public posting of summary results within a certain time frame (15), represent one type of response. However, improving transparency is only part of the solution to the broader set of concerns about medical interventions.

## Promoting Transparency

Transparency exists along a continuum from documentation that a trial exists to full disclosure of the results data set at the end of the trial (see figure, below). “Trial registries” address one end of the spectrum by making public a summary of protocol details at trial initiation.

As new policies promote transparency of clinical trials through registries and results databases, further issues arise and require examination.

rate of trial registration has stayed constant at about 250 new trials per week, resulting in nearly 44,800 trials from 150 countries as of September 2007.

Since 2005, the U.S. Congress has considered bills calling for clinical trial registration and results reporting (22) and, on 27 September 2007, enacted the FDA Amendments Act (FDAAA) (23). Section 801 of this law (“FDAAA 801”) expands the scope of required registrations at ClinicalTrials.gov and provides for the first federally funded trial results database. It mandates registration of a set of controlled clinical investigations, other than phase I trials, of drugs, biologics, and devices subject to regulation by the FDA. The law applies to research for any condition regardless of sponsor type (e.g., industry, government, or academic). These new statutory requirements, although broader than previous law, remain narrower than the transnational policies of the ICMJE and the World Health Organization (WHO), which call for the registration of all interventional studies in human beings regardless of intervention type (11, 24). FDAAA 801 also increases the number of mandatory data elements, corresponding to the WHO and ICMJE international standard, and requires ClinicalTrials.gov to link the registry to specified, existing results information publicly available from the FDA Web site, including summary safety and effectiveness data, public health advisories, and action packages for drug approval.

Numerous clinical trial registries exist; the NIH has maintained ClinicalTrials.gov, the largest single registry of clinical trials, since 2000 (18). Although the law that led to the creation of ClinicalTrials.gov, the Food and Drug Administration Modernization Act of 1997, called for the registration of some trials of drug effectiveness for “serious or life-threatening diseases and conditions” (19), other registration policies have encouraged broader voluntary registration of trial information. A policy by the International Committee of Medical Journal Editors (ICMJE) that requires prospective trial registration as a precondition for publication, effective September 2005 (20), led to a 73% increase in trial registrations of all intervention types from around the world (21). This increased

“Results databases” provide public summaries of results for key trial end points, whether published or not. Some policies promote public access to data sets, such as the National Institutes of Health (NIH) Data Sharing Policy (16) and the policy of the *Annals of Internal Medicine* policy, which publishes author statements of willingness to share study protocols, statistical codes, and data sets (17).

Numerous clinical trial registries exist; the NIH has maintained ClinicalTrials.gov, the largest single registry of clinical trials, since 2000 (18). Although the law that led to the creation of ClinicalTrials.gov, the Food and Drug Administration Modernization Act of 1997, called for the registration of some trials of drug effectiveness for “serious or life-threatening diseases and conditions” (19), other registration policies have encouraged broader voluntary registration of trial information. A policy by the International Committee of Medical Journal Editors (ICMJE) that requires prospective trial registration as a precondition for publication, effective September 2005 (20), led to a 73% increase in trial registrations of all intervention types from around the world (21). This increased



National Library of Medicine, National Institutes of Health, U.S. Department of Health and Human Services, Bethesda, MD 20894, USA.

\*Author for correspondence. E-mail: [dzarin@mail.nih.gov](mailto:dzarin@mail.nih.gov)

### Will FDAAA 801 Solve Recent Problems?

The table (below) illustrates a typology of public concerns about the system of evaluating drugs and devices. We have categorized selected recent controversial issues by alleged problem: design, conduct, or analysis of the study; lack of public information about the study existence or results; and regulatory agency decision-making.

FDAAA 801 directly addresses issues stemming from a lack of transparency in clinical trials, represented by the examples in the highlighted section of the table. For instance, GlaxoSmithKline (GSK)-sponsored trial data for the heavily promoted antidepressant Paxil showing efficacy and safety concerns in children and adolescents were not available to the public (25, 26). The resulting 2004 legal settlement between GSK and the New York State Attorney General's office required GSK to develop a publicly accessible, online results database (27) for the timely, comprehensive posting of trial results for company-marketed drugs (28, 29). In the case of Vioxx, a COX-2 nonsteroidal anti-inflammatory drug (NSAID), "Several early, large clinical trials ... were not published in the academic literature for years after Merck made them available to the FDA, preventing independent investigators from accurately determining its cardiovascular risk..." (p. 122) (30). With evidence that Vioxx is associated with increased cardiovascular risk from subsequent clinical trials, the drug was voluntarily withdrawn from the market in September 2004.

More recently, questions have focused on the ENHANCE trial, an industry-sponsored study of Zetia, a marketed nonstatin cholesterol-lowering drug that is also a component of the cholesterol-lowering drug Vytorin. Issues include delayed trial registration and results reporting, as well as attempting to modify prespecified primary outcome measure (31). Results, generally regarded as negative, were revealed in a company press release (32) after intense media attention and a congressional investigation (2).

Issues related to the design or conduct of clinical trials, including research ethics, are not covered by FDAAA 801. For example, allegations of human research protections violations in a 1996 Trovan (antibiotic) study on children in Zaire (33) and data integrity questions about Ketek (antibiotic) study 3014, in which FDA inspectors detected data fraud and other serious violations (34), would not be affected by FDAAA 801. Further, Merck studies of Arcoxia, a COX-2 NSAID, involving over 34,000 patients, were judged to be of limited scientific interest because of the use of

an "inappropriate" comparator with many known side effects of its own (35). Nevertheless, it is possible that complete trial registration and results reporting might have helped institutional review boards (IRBs) assess the need for each additional Arcoxia trial.

Although current policies have focused on interventional studies, observational studies play an increasingly critical role in biomedical research, especially in the assessment of safety after an intervention comes into widespread use. Postmarket observational studies provide data about rare, unanticipated adverse effects from the exposure of large numbers of heterogeneous individuals over periods of time longer than typically studied in controlled trials (36). Despite methodological limitations, such as susceptibility to confounding factors and other sources of bias, which potentially lead to inconclusive or misleading results [e.g., data on risks of postmenopausal hormone replacement therapy from the Women's Health Initiative (37)], observational research, nevertheless, can play a useful, complementary role to interventional studies (38). Yet observational studies have received less attention in the quest for transparency.

In the case of a cholesterol-lowering drug (Baycol), a company-conducted observational study showed a higher relative rate of muscle breakdown for Baycol compared with another marketed statin. This finding was never reported publicly, but became available as a result of litigation (39). In another case, results of a Bayer-commis-

sioned postmarket observational study of 67,000 patients that raised concerns about associations between Trasyolol, a clotting drug, and cardiovascular or renal risk were not released by the company until after an FDA advisory committee meeting to evaluate the safety of Trasyolol (36, 40). In November 2007, Bayer announced the voluntary suspension of Trasyolol sales worldwide, pending further analysis of safety data. Some cases reflect a deficiency in monitoring serious adverse events once a drug or device is marketed. In other cases, such as Guidant's Ventak Prizm 2 DR implantable cardioverter-defibrillator, information about a problematic adverse event profile (i.e., device malfunctions) was not publicly disseminated in a timely fashion (41).

Other concerns relate to thresholds for determining what and how much data are sufficient to prompt regulatory decisions regarding the availability or labeling of a medical product. For example, questions have been raised about the timeliness and adequacy of FDA's response to data on Avandia (42), trials of Ketek (34), and on the use of some antidepressants in children (43, 44). Transparency policies alone do not address these issues, although they could help to empower members of the public who believe that regulatory action is warranted. For example, FDA mandated changes to the Avandia label in November 2007 to reflect the risk of myocardial infarction; this risk was first publicized in a meta-analysis published in June 2007 (3) that used, in part, data from the GSK database mandated by the Paxil settlement.

CATEGORY OF PERCEIVED PROBLEM	SAMPLE ISSUES	RECENT EXAMPLES	ADDRESSED BY § 801?
<b>Design, conduct, or analysis</b>	<ul style="list-style-type: none"> <li>• Appropriateness of comparator</li> <li>• Lack of data integrity/fraud</li> <li>• Insufficient informed consent</li> </ul>	Arcoxia (35) Ketek (34) Trovan (33)	No
<b>Lack of public information</b>			
Clinical trial	<ul style="list-style-type: none"> <li>• Suppression of trial existence and results</li> </ul>	Paxil (40) Vioxx (30) Zetia (31)	Yes
Observational study	<ul style="list-style-type: none"> <li>• Suppression of study existence and results withheld</li> </ul>	Baycol (39) Trasyolol (36)	No
Postmarket adverse event reports	<ul style="list-style-type: none"> <li>• Failure to disseminate data</li> </ul>	ICD* (41)	Pending†
<b>Regulatory agency decision-making</b>	<ul style="list-style-type: none"> <li>• Delayed agency disclosure</li> <li>• Delayed agency action</li> </ul>	Ketek (34) Avandia (42)	No

\*Ventak Prizm 2 DR implantable cardioverter-defibrillator. †For example, postmarket surveillance is addressed in FDAAA, § 905.

Sample medical product safety concerns by category of perceived problems.

### Future Challenges

FDAAA 801 is intended to greatly expand the level of transparency for clinical trials, which could have a transforming effect on our system of evaluating drugs and devices. However, FDAAA 801 still leaves areas of "opacity." For example, although certain medical device trials must be registered at trial inception, this information is withheld from the public until device approval or clearance by FDA. The resulting "lockbox" prevents disclosure of trials of devices that are never approved or cleared (e.g., where safety concerns arose or sponsors abandoned further development). For instance, Boston Scientific stopped development of an experimental stent after clinical trials revealed frequent fractures in the device (45). In addition FDAAA 801 does not mandate public reporting for phase I drug trials and trials involving investigational interventions not regulated by the FDA, such as surgical procedures and behavioral therapies. Thus, lessons from phase I trials, such as the life-threatening adverse events in healthy volunteers caused by the superagonist, TGN1412, could go unreported to the public and, potentially, result in redundant studies by future unsuspecting researchers (46). Further, results reporting is currently mandatory only for trials of FDA-approved medical products, allowing the results of unapproved products to remain hidden from public view.

**Intellectual property-related issues.** Current restrictions reflect a delicate balance between protection of commercial interests and promotion of public health (15). Pharmaceutical, biotech, and medical device manufacturers are concerned that disclosures may undermine competitive advantage (47). However, there are important ethical and scientific reasons for broader disclosure: Trial participation by humans is predicated on the concept that the trial will add to "medical knowledge," which requires dissemination of the results. In addition, it is not possible for a volunteer or an IRB to assess the risks and benefits of participation in a clinical trial if an unknown proportion of data on the proposed interventions is not publicly available (48). FDAAA 801 calls on the Secretary of Health and Human Services to consider whether public health interests would support expanding the results requirements to include unapproved drugs, biologics, and devices, through a 3-year rule-making process.

**Validation of information.** Concerns have been raised about verifying the completeness or accuracy of sponsor- and researcher-submitted results data. The vol-

ume of completed trials, the lack of access to protocols and data sets, and the subjective nature of some judgments are barriers to validation. The law reflects these concerns and mandates the reporting of objective data in tables. The NIH and FDA are directed by the law to conduct a pilot quality-control project to determine the optimal method of verification. In addition, narrative summaries could be required in the future, but only if "such types of summaries can be included without being misleading or promotional" (23). Additional research will be necessary to explore whether and how this might be accomplished.

**Interpretation of information.** The results database will require an interface to assist users in finding study results. One concern is that members of the lay public and media may be ill-prepared to interpret summary results data (49). Currently, there are no standards or guidelines for providing and explaining study results to trial participants or to members of the public (50). FDAAA 801 calls for the development of informational materials for consumers in consultation with health and risk communication experts. Furthermore, clinicians may be concerned that the existence of a results database will increase the patient expectations that clinicians will be knowledgeable about all results in the database, even those that were never published or discussed in a journal. In addition, although the results database will facilitate the conduct of carefully performed and comprehensive systematic reviews, some also worry that the public will have a difficult time assessing the quality of the multitude of analyses that may result (51).

### Conclusion

FDAAA 801 should transform the degree of public access to critical clinical trial information from publicly and privately funded clinical research. J. M. Drazen, Editor-in-Chief, *New England Journal of Medicine*, has noted that currently, some patients are "left on the cutting room floor to make a drug look better than it really is" (52); FDAAA 801 should go a long way in ensuring that all patients and all data are publicly accounted for.

### References and Notes:

1. E. H. Turner, A. M. Matthews, E. Linardatos, R. A. Tell, R. Rosenthal, *N. Engl. J. Med.* **358**, 252 (2008).
2. A. Berenson, *New York Times*, 15 January 2008.
3. S. E. Nissen, K. Wolski, *N. Engl. J. Med.* **356**, 2457 (2007).
4. R. Stein, *Washington Post*, 22 May 2007, p. A3.
5. G. Harris, *New York Times*, 12 September 2007.
6. K. Dixon, *Reuters*, 26 June 2007.
7. Editorial *Lancet* **369**, 1834 (2007).
8. C. J. Rosen, *N. Engl. J. Med.* **357**, 844 (2007).

9. "Rosiglitazone: Seeking a balanced perspective," *New York Times*, 6 June 2004.
10. J. M. Drazen, S. Morrissey, G. D. Curfman, *N. Engl. J. Med.* **357**, 1756 (2007).
11. C. Laine et al., *N. Engl. J. Med.* **356**, 2734 (2007).
12. R. J. Simes, *J. Clin. Oncol.* **4**, 1529 (1986).
13. P. J. Easterbrook, J. A. Berlin, R. Gopalan, D. R. Matthews, *Lancet* **337**, 867 (1991).
14. R. T. Johnson, K. Dickerson, *Nat. Clin. Pract. Neurol.* **3**, 590 (2007).
15. C. B. Fisher, *Science* **311**, 180 (2006).
16. U.S. National Institutes of Health, "NIH Data Sharing Policy and Implementation Guidance" (NIH, Bethesda, MD, 2003); [http://grants.nih.gov/grants/policy/data\\_sharing/data\\_sharing\\_guidance.htm](http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm).
17. Information for Authors, *Ann. Intern. Med.*; [www.annals.org/shared/author\\_info.html](http://www.annals.org/shared/author_info.html).
18. D. A. Zarin et al., *JAMA* **297**, 2112 (2007).
19. Food and Drug Administration Modernization Act of 1997, Public Law No. 105-115 § 113 (1997).
20. C. D. DeAngelis et al., *JAMA* **292**, 1363 (2004).
21. D. A. Zarin, T. Tse, N. C. Ide, *N. Engl. J. Med.* **353**, 2779 (2005).
22. E. D. Williams, "Clinical Trials Reporting and Publication," July 2007 (Congressional Research Service, Washington, DC, 2007).
23. FDAAA, Public Law No. 110-85 § 801 (2007).
24. I. Sim et al., *Lancet* **367**, 1631 (2006).
25. D. Rennie, *JAMA* **292**, 1359 (2004).
26. *The People of the State of New York v. GlaxoSmithKline*, Complaint, filed 2 June 2004.
27. GlaxoSmithKline Clinical Trial Register; <http://ctr.gsk.co.uk/welcome.asp>.
28. *The People of the State of New York v. GlaxoSmithKline*, Consent Order & Judgment: Civil Action No. 04-CV-5304 MGC, ordered 26 August 2004.
29. O. Dyer, *BMJ* **329**, 590 (2004).
30. H. M. Krumholz, J. S. Ross, A. H. Presler, D. S. Egilman, *BMJ* **334**, 120 (2007).
31. A. Berenson, *New York Times*, 21 December 2007.
32. "Merck/Schering-Plough Pharmaceuticals provides results of the ENHANCE trial," Product News, 14 January 2008 (Merck & Co., Whitehouse Station, NJ, 2008); [www.merck.com/newsroom/press\\_releases/product/2008\\_0114.html](http://www.merck.com/newsroom/press_releases/product/2008_0114.html).
33. J. Stephens, *Washington Post*, 30 May 2007, p. A10.
34. D. B. Ross, *N. Engl. J. Med.* **356**, 1601 (2007).
35. A. Berenson, *New York Times*, 24 August 2006.
36. W. R. Hiatt, *N. Engl. J. Med.* **355**, 2171 (2006).
37. C. Laine, *Ann. Intern. Med.* **137**, 290 (2002).
38. J. Avorn, *N. Engl. J. Med.* **357**, 2219 (2007).
39. B. M. Psaty, C. D. Furberg, W. A. Ray, N. S. Weiss, *JAMA* **292**, 2622 (2004).
40. J. Avorn, *N. Engl. J. Med.* **355**, 2169 (2006).
41. R. G. Hauser, B. J. Maron, *Circulation* **112**, 2040 (2005).
42. B. M. Psaty, C. D. Furberg, *N. Engl. J. Med.* **357**, 67 (2007).
43. R. Waters, *San Francisco Chronicle*, 1 February 2004, p. A1.
44. C. Holden, *Science* **303**, 745 (2004).
45. B. Meier, *New York Times*, 30 October 2007.
46. M. Goodyear, *BMJ* **332**, 677 (2006).
47. H. I. Miller, D. R. Henderson, *Nat. Rev. Drug Discov.* **6**, 532 (2007).
48. L. A. Levin, J. G. Palmer, *Arch. Intern. Med.* **167**, 1576 (2007).
49. A. J. Pinching, *J. R. Soc. Med.* **88** (Suppl. 24), 12 (1995).
50. A. H. Partridge, E. P. Winer, *JAMA* **288**, 363 (2002).
51. E. Wager, *PLoS Clin. Trials* **1**, e31 (2006).
52. A. W. Matthews, A. Johnson, *Wall Street Journal*, 23 January 2008, p. A14.
53. Supported by the Intramural Research Program of the NIH, National Library of Medicine. We thank J. Sheehan for comments on the draft manuscript. The ideas and opinions expressed are the authors'. They do not represent any policy position of the NIH, Public Health Service, or Department of Health and Human Services.

## GEOLOGY

# Canyon Cutting on a Grand Time Scale

Tim Atkinson<sup>1</sup> and Mike Leeder<sup>2</sup>

When Hutton first conceived the immensity of geologic time, he remarked that he could see in the record of the rocks “no vestige of a beginning, no prospect of an end” to the upheaval and wearing down of mountains (1). Nineteenth-century geologists elaborated the notion that continuous erosion produced debris carried by rivers, glaciers, and wind to lowlands and the sea, where sediments accumulated to eventually form rock strata. Hutton’s crucial observation was that some of these strata were tilted by upheavals of the crust, only to be worn down again by erosion in renewed cycles of immense duration.

The question of just how immense fell to 20th-century scientists to answer. Through a plethora of techniques for measuring rock ages, we now understand Earth’s sedimentary history in remarkable detail. Also understood, although in much broader terms, are the episodes of crustal uplift, which are caused by vertical motions ancillary to the horizontal motions of tectonic plates. But interpreting the ages and histories of continental land surfaces has lagged far behind. On page 1377 of this issue, Polyak *et al.* (2) elucidate this side of the cycle. They show that the Colorado River has taken ~20 million years to incise its course downward through the 1-mile depth of the Grand Canyon. This time scale is not surprising—many geologists have long suspected it—but the study uses an ingenious combination of methods to demonstrate it firmly for the first time.

Hutton and Lyell might have thought 20 million years an immensely long time, but it is very short relative to the 4500-million-year age of Earth, and short even compared with the ages of the strata through which the Grand Canyon has been cut. The new age for such a major valley confirms that on geological time scales, the relief of the continents is continually remade by the interplay of denudation and tectonic forces.

The earliest estimates of the rate of denudation of the continents were made in the 19th century on the basis of sparse data on water



**An artist's depiction of the Grand Canyon.** In 1879, the British geologist Archibald Geikie met J. W. Powell and C. E. Dutton, pioneers of western U.S. geology, in Utah. He used the original black-and-white version of this drawing by U.S. Geological Survey geologist-artist W. H. Holmes as frontispiece for his influential *Text-Book of Geology* (3).

discharge and sediment concentrations in major rivers. Attempts to estimate the rate of incision of valleys go back at least to Archibald Geikie (3). His value of 1 foot (31 cm) in 1200 years is in the mid-range of the new data for the Grand Canyon, but it was merely an educated guess; Geikie realized that he lacked most of the information essential for assessing the incision rate of any valley accurately. Because erosion continually wears almost all topography away, some fortuitously preserved benchmark is needed that can indicate the level of the valley floor at a certain time in the past, and there must be a means of measuring the age of each and every benchmark. Comparing the heights and ages of the benchmarks with the present topography then allows the rate of valley incision to be deduced.

This is the method followed by Polyak *et al.*, who use cave deposits marking former positions of the water table in the Grand Canyon’s walls as benchmarks of ancient levels of the river bed. Their age-dating method exploits the radioactive decay of uranium via other elements to lead. In common with others (4, 5), they have used recent advances in mass spectrometry to extend the younger limit of this technique to a few hundred thousand years and

New data show how the Grand Canyon has been formed over the course of the past 20 million years.

apply it to cave deposits. All previous uses of cave deposits for landform-dating were hampered by lack of a technique with an age limit exceeding 1 to 3 million years (6–8). This limitation is removed by the new adaptation of the uranium-lead (U-Pb) method. The study also elucidates the role tectonics played in stimulating the incision of the great system of canyons. Cutting a deep gorge through an elevated plateau requires a through-flowing river and progressive development of a topographic step to cause erosion to propagate upstream. Polyak *et al.* identify a spatial migration of canyon formation and a change in incision rate, with older, slower down-cutting in the west and younger, faster down-cutting in the east.

Previously, incision rates were extrapolated from radiometric dating of young lava flows preserved on low rock terraces throughout the canyon system (9). These rates were deemed too small to have cut the whole canyon in the ~6 million years required by some indirect geological evidence for the initiation of the through-flowing Colorado River: Either the rates were higher further back in time, or some canyon relief had been cut earlier. The new study supports the latter alternative (10). The

<sup>1</sup>Department of Earth Sciences, University College London, London WC1E 6BT, UK. E-mail: t.atkinson@ucl.ac.uk

<sup>2</sup>School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK. E-mail: m.leeder@uea.ac.uk

necessary western topographic step must have been put in place by tectonics about 20 million years ago. This arose by a combination of broad regional uplift due to upper mantle processes (11) and initiation of the Grand Wash Fault at the western edge of the Colorado Plateau. The Colorado River of 20 million years ago poured over this growing step on its way westward.

Each region of the continental crust is uplifted episodically by tectonic forces, but it also adjusts continuously to the unloading caused by erosion itself, rising at a rate slightly less than that of regional denudation [the average rate at which rock is stripped from the

region as a whole (7)]. The current average denudation rate for the drainage area above Lake Mead is 145 to 160 m per million years (12, 13); on the Colorado Plateau itself, it has been measured as 26 m per million years (13). The incision rate of 166 to 411 m per million years found by Polyak *et al.* in the eastern Grand Canyon exceeds both these estimates. This means that uplift there was too fast to have been driven by erosion alone. Tectonic forces must have contributed in the past and possibly continue to do so today.

#### References

1. J. Hutton, *Trans. R. Soc. Edinburgh* **1**, 209 (1788).
2. V. Polyak, C. Hill, Y. Asmerom, *Science* **319**, 1377 (2008).

3. A. Geikie, *Text-Book of Geology* (Macmillan, London, 1882), pp. 441–447.
4. D. A. Richards *et al.*, *Geochim. Cosmochim. Acta* **62**, 3683 (1998).
5. J. Walker *et al.*, *Science* **314**, 1592 (2006).
6. D. C. Ford *et al.*, *Arctic Alpine Res.* **13**, 1 (1981).
7. A. R. Farrant *et al.*, *Geology* **23**, 357 (1995).
8. T. C. Atkinson, P. J. Rowe, in *Uranium-Series Disequilibrium; Applications to Earth, Marine, and Environmental Sciences*, M. Ivanovich, R. Harmon, Eds. (Oxford Univ. Press, New York, 1992), pp. 669–703.
9. J. Pederson *et al.*, *Geology* **30**, 739 (2002).
10. R. F. Holm, *Bull. Geol. Soc. Am.* **113**, 1467 (2001).
11. M. Roy *et al.*, *Geology* **32**, 925 (2004).
12. W. O. Smith *et al.*, *Prof. Pap. U.S. Geol. Surv.* 295 (1961).
13. L. B. Leopold *et al.*, *Fluvial Processes in Geomorphology* (Freeman, San Francisco, 1964), p. 76.

10.1126/science.1155286

## MOLECULAR BIOLOGY

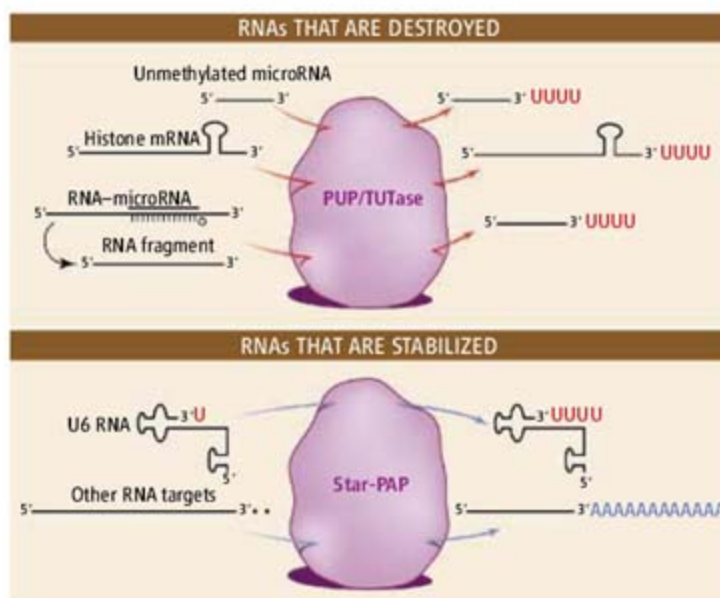
# A Tail Tale for U

Marvin Wickens and Jae Eun Kwak

Nothing we learn about the ribonucleic acid (RNA) world should surprise us any more. RNAs that are destroyed just after synthesis, or salvaged from cellular scrap heaps, or that regulate transcription—the process by which RNAs themselves are made—already are commonplace. Yet surely a few things in this fluid world are fixed, such as the string of adenosine bases [poly(A) tail] that is added to the end of eukaryotic messenger RNAs (mRNAs), the templates transcribed from DNA that encode proteins. Now even this terra firma is unstable. Recent work identifies enzymes that synthesize polyuridine [poly(U)], and shows that poly(U) tails not only exist on certain mRNAs, but can control mRNA decay.

Many RNAs receive extra nucleotides at their 3' ends after transcription. These “tails” are not copied from DNA, but are synthesized *de novo*. Poly(A) is the best known, contributing to the stability and export of mRNAs and their translation to proteins. However, U's are added to the ends of guide RNAs that function in RNA editing (1) and to certain microRNAs (2) and small nuclear RNAs (3).

Several enzymes that catalyze “tailing” belong to a superfamily of DNA polymerase  $\beta$ -like nucleotidyl transferases (4).



**Dual-personality enzyme.** Nucleotide polymerases can add a poly(U) tail (PUP/TUTase activity) or poly(A) tail (Star-PAP activity) to specific RNAs to control their degradation or stability.

This includes poly(A) polymerases (PAPs), which add poly(A) to mRNAs. Recently, a new family of enzymes that add poly(U) was identified (5). These poly(U) polymerases (PUPs) are widespread, from yeast to humans. Another member of the same superfamily called terminal uridylyl transferase (TUTase) adds short U tails [oligo(U)] to a noncoding RNA, U6 (1, 6). Although PUPs and TUTases diverge substantially in sequence, their functions may overlap.

A new PUP-related enzyme identified by Mellman *et al.* (7) is particularly provocative because it adds U's to one type of RNA and A's to others. Initially, it was shown that the mammalian enzyme U6 TUTase adds one to three U's

Enzymes add strings of uridine or adenosine to control RNA stability.

to the end of U6 RNA, a small RNA involved in RNA splicing (6). Mellman *et al.* show that the same enzyme is a divergent PAP, dubbed Star-PAP. Star-PAP is nuclear, and associates with proteins that cleave the 3' end of precursor mRNAs, implying an alternative mRNA processing machine. Candidate mRNA targets have been identified and include many involved in the oxidative stress response.

Star-PAP also binds to phosphatidylinositol 4-phosphate 5-kinase  $\alpha$ , an enzyme that generates a specific phosphoinositide, phosphatidylinositol 4,5-bisphosphate. Remarkably, the same phosphoinositide stimulates Star-PAP activity. Thus, perhaps in response to stress, the kinase produces a small-molecule signal (a phosphoinositide) that specifies the maturation of targets mRNAs. Another enzyme that targets RNA, ADAR2, is regulated by a different phosphoinositide (8). Very likely, small molecules regulate other enzymes that act on RNA, but have escaped notice.

Like Star-PAP, other PUPs may be dual-purpose. For example, in the yeast *Schizosaccharomyces pombe*, CID1 can add either U's or A's to RNAs *in vitro* (5, 9). Although CID1 and other PUPs synthesize only poly(U) in the cell (5), their nucleotide preference could be modulated by small molecules or vary with the targeted RNA. With Star-PAP, the phosphoinositide could trigger such a switch in activity. *In vivo* studies are needed to test this notion decisively, though detecting RNAs that have received poly(U)

Department of Biochemistry, University of Wisconsin, Madison, WI 53706, USA. E-mail: [wickens@biochem.wisc.edu](mailto:wickens@biochem.wisc.edu)

will be challenging if they are unstable.

New functions for poly(U) are emerging. The tail can be added to mRNAs to trigger their decay. mRNAs that direct the synthesis of histones—the major protein constituents of chromatin—during the cell division cycle are rapidly degraded once DNA replication is completed or blocked. Intermediate mRNAs in this decay process often have poly(U) tails of 8 to 10 residues (10). Small interfering RNAs directed against two candidate PUP enzymes blocked this degradation, implying that poly(U) addition is essential for their decay. Also, in *S. pombe*, a PUP adds poly(U) to actin mRNA, though its effect on turnover is unknown (9).

A poly(U) tail may enhance degradation by stimulating removal of the mRNA's 5' cap structure, a key step in mRNA turnover. Poly(U) tails enhance "decapping" in a cell-free system (11). Likely, the tails bind the Lsm protein complex, which associates with decap-

ping factors (11, 12). Indeed, depletion of Lsm1 inhibits histone mRNA turnover (10).

Addition of uridines probably has diverse consequences, including RNA stabilization (6); yet this modification often occurs on an RNA's road to ruin (see the figure). Aberrantly unmethylated microRNAs in the plant *Arabidopsis thaliana* are modified with oligo(U) and destroyed (2). Fragmentation of mRNA by microRNAs is accompanied by the addition of oligo(U) to the pieces before they disappear (13). The most common mRNA decay pathway involves association of the Lsm complex to the mRNA after poly(A) removal. Even this route may rely on evanescent, short oligo(U) because the Lsm complex preferentially binds 3'-terminal uridine tails.

The discovery of poly(U) tails on mRNAs opens unexplored territory in the RNA world. Dual-personality enzymes could switch an mRNA's fate from life to death simply by a

change in the nucleotide they accept. Others may well wait in the wings, along with proteins that target specific RNAs, or remove the tails. Count on new roles for poly(U), an expanding list of RNAs that receive it, and more startling enzymes that put it on and take it off.

#### References and Notes

1. R. Aphasizhev, *Cell. Mol. Life Sci.* **62**, 2194 (2005).
2. J. Li, Z. Yang *et al.*, *Curr. Biol.* **15**, 1501 (2005).
3. E. Lund, J. E. Dahlberg, *Science* **255**, 327 (1992).
4. G. Martin, W. Keller, *RNA* **13**, 1834 (2007).
5. J. E. Kwak, M. Wickens, *RNA* **13**, 860 (2007).
6. R. Trippe *et al.*, *RNA* **12**, 1494 (2006).
7. D. L. Mellman *et al.*, *Nature* **451**, 1013 (2008).
8. M. R. Macbeth *et al.*, *Science* **309**, 1534 (2005).
9. O. S. Rissland, A. Mikulasova, C. J. Norbury, *Mol. Cell. Biol.* **27**, 3612 (2007).
10. T. E. Mullen, W. F. Marzluff, *Genes Dev.* **22**, 50 (2008).
11. M. G. Song, M. Kiledjian, *RNA* **13**, 2356 (2007).
12. S. Tharun *et al.*, *Nature* **404**, 515 (2000).
13. B. Shen, H. M. Goodman, *Science* **306**, 997 (2004).
14. We thank R. Parker, J. Kimble, and the Wickens lab for comments. Work in the Wickens lab is supported by the NIH.

10.1126/science.1154946

## BEHAVIOR

# Punishment and Cooperation

Herbert Gintis

Even champions of modern society agree that it involves a loss of community (based on family and ethnic ties) and an expansion of civil society, with emphasis on the more impersonal interactions among individuals with minimal social ties. For two centuries, this dichotomy has anchored our understanding of modern Western society, applauded by its defenders as the fount of freedom (1), yet identified as the source of inequality (2), the decline of community (3), the destruction of the environment (4), and the impotence of grassroots political action (5). On page 1362 of this issue, Herrmann *et al.* (6) report their discovery that university students in democratic societies with advanced market economies show different social behavior from that exhibited by students in more traditional societies based on authoritarian and parochial social institutions. Their results suggest that the success of democratic market societies may depend critically on moral virtues as well as material interests, so the depiction of civil society as the sphere of "naked self-interest" is radically incorrect.

The standard view holds that human nature has a private side in which we interact morally with a small circle of intimates and a public

side in which we behave as selfish maximizers. Herrmann *et al.* suggest that most individuals have a deep reservoir of behaviors and mores that can be exhibited in the most impersonal interactions with unrelated others. This reservoir of moral predispositions is based on an innate prosociality that is a product of our evolution as a species, as well as the uniquely human capacity to internalize norms of social behavior. Both forces predispose individuals to behave morally even when this conflicts with their material interests.

These results are the latest to document a principle of reciprocity according to which people are more willing to sacrifice private gain for the public good as the cost of the sacrifice decreases and as expectations of the extent that others will sacrifice grows. In addition, individuals embrace such character virtues as honesty, trustworthiness, consideration, and loyalty (7). Of course, these moral predispositions moderate rather than eliminate considerations of self-interest and loyalties to kith and kin.

Suggestive evidence for the principle of reciprocity comes from daily life. For instance, political democracy has frequently been attained through popular collective action. Voting in elections is widespread despite its being personally time consuming, and the

Data from economic games show that the effectiveness of punishment in fostering cooperation varies greatly from society to society.

benefits are purely public (a single vote can change an electoral outcome only with infinitesimal probability). Moreover, citizens in democratic societies often vote to give substantial sums to charity, and to approve of poverty relief, although these measures increase the tax burden for the average voter.



Experimental evidence for reciprocity comes from behavioral game theory, which uses economic games in which subjects make choices under varied social conditions. For instance, Herrmann *et al.* employ a public goods game in which each of four anonymous subjects is initially given 20 tokens, and each is told he can place any number of these tokens in a public account. The tokens in the account are multiplied by 1.6 and the result divided evenly among the four. At the end of the experiment, the tokens are exchanged for real money.

In this game, each individual helps the group most by placing his 20 tokens in the

public account, and if all do so, each earns 32 tokens. However, if a single individual is selfish, he will place nothing in the public account, and his earnings will be  $20 + 60(1.6)/4 = 44$  tokens. But, if all four are selfish, each earns only 20 tokens. Because the four subjects are strangers, the standard view of human nature suggests that there will be zero contributions. However, in the many times this game has been played in a variety of social settings, the older view is virtually never supported, and the average contribution is about half the initial endowment (8).

The public game indicates that individuals generally fall halfway between selfishness (keep all 20 tokens) and public-spiritedness (place all 20 tokens in the public account). However, mean contributions to the public account generally fall over many trials, reaching a very low level after 10 repetitions. By varying the rules of the game, researchers have concluded that the principle of reciprocity is responsible for the observed decay of cooperation: Subjects who contributed more than average on one round contribute less on the following round, showing their disapproval of the unfairness of their fellow players. Indeed, a single selfish individual in the group can lead contributions to spiral down to almost zero.

An innovation of Fehr and Gächter (9), used by Herrmann *et al.* as well, was to add punishment after each round of play. Each player A could specify that the player B associated with a particular contribution have three tokens deducted from his payoff, for each token deducted from A's payoff. Under these new conditions, the high contributors punished the low contributors who, in succeeding rounds, increased their contributions, so that in the 10th and final round, there was almost 100% cooperation. The behavioral propensity to cooperate with others at personal cost, and to punish non-cooperators even when this is personally costly in the long run, has been called strong reciprocity. The punishment meted out is considered altruistic because it increases the payoff of group members at a personal cost to the punisher.

The natural interpretation is that low contributors are selfish types who increase their contribution after punishment in order to avoid future punishment. However, many low contributors respond almost as much to symbolic as to monetary penalties, which indicates that many punishers are not self-interested but rather are motivated to increase their offers because they feel guilty for having violated a contribution norm (10). However, some researchers found a curious phenomenon. A few subjects, when punished, rather than contributing more,

suspected that it was the high contributors who punished them, and responded with antisocial punishment: They punished the high contributors in future rounds, leading the latter to reduce both their contribution and altruistic punishment (11).

Herrmann *et al.* collected data in 15 countries with widely varying levels of economic development. The subjects were university students in all societies. The authors found that antisocial punishment was rare in the most democratic societies and very common otherwise. Indexed to the World Democracy Audit (WDA) evaluation of countries' performance in political rights, civil liberties, press freedom, and corruption, the top six performers among the countries studied were also in the lowest seven for antisocial punishment. These were the United States, the United Kingdom, Germany, Denmark, Australia, and Switzerland. The seventh country in the low antisocial punishment group was China, currently among the fastest-growing market economies in the world. The countries with a high level of antisocial punishment and a low score on the WDA evaluation included Oman, Saudi Arabia, Greece, Russia, Turkey, and Belarus.

The most likely explanation is that in more traditional societies, the experimental setup represents a clash of cultures. On the one hand, high payoffs in the experiment require the modern

ethic of cooperation with unrelated strangers, so subjects who are reprimanded for low contribution are likely to respond with feelings of guilt and a resolve to be more cooperative in the future. In a more traditional society, many players may hold to the ethic of altruism and sacrifice on behalf of one's family and friends, with indifference toward unrelated strangers. When punished, such subjects are likely to respond with anger rather than guilt. Punishing the high contributors is thus a means of asserting one's personal values, which take precedence over maximizing one's payoff in the game.

#### References and Notes

1. M. Friedman, *Capitalism and Freedom* (Univ. of Chicago Press, Chicago, 1962).
2. M. D. Young, *The Rise of the Meritocracy* (Thames and Hudson, London, 1958).
3. W. C. McWilliams, *The Idea of Fraternity in America* (Univ. of California Press, Berkeley, CA, 1973).
4. G. Hardin, *Science* **162**, 1243 (1968).
5. M. Olson, *The Logic of Collective Action: Public Goods and the Theory of Groups* (Harvard Univ. Press, Cambridge, MA, 1965).
6. B. Herrmann *et al.*, *Science* **319**, 1362 (2008).
7. U. Gneezy, *Am. Econ. Rev.* **95**, 384 (2005).
8. J. O. Ledyard, in *The Handbook of Experimental Economics*, J. H. Kagel, A. E. Roth, Eds. (Princeton Univ. Press, Princeton, NJ, 1995), pp. 111–194.
9. E. Fehr, S. Gächter, *Nature* **415**, 137 (2002).
10. D. Masclot *et al.*, *Am. Econ. Rev.* **93**, 366 (2003).
11. O. Bochet *et al.*, *J. Econ. Behav. Organ.* **60**, 11 (2006).
12. I thank Samuel Bowles for advice in the preparation of this article.

10.1126/science.1155333

## OCEANS

# On Phytoplankton Trends

Victor Smetacek<sup>1</sup> and James E. Cloern<sup>2</sup>

How are phytoplankton at coastal sites around the world responding to ongoing global change?

Phytoplankton—unicellular algae in the surface layer of lakes and oceans—fuel the lacustrine and marine food chains and play a key role in regulating atmospheric carbon dioxide concentrations. How will rising carbon dioxide concentrations in the air and surface ocean in turn affect phytoplankton? Answering this question is crucial for projecting future climate change. However, because phytoplankton species populations appear and disappear within weeks, assessing change requires high-resolution monitoring

of annual cycles over many years. Such long-term studies at coastal sites ranging from estuaries and harbors to open coastlines and islands are yielding bewildering variability, but also fundamental insights on the driving forces that underlie phytoplankton cycles (1).

An example of regularity is provided by a 45-year data set from weekly phytoplankton monitoring in Lake Windermere, England, which shows that the diatom species *Asterionella formosa* dominates phytoplankton biomass from autumn to spring but is virtually absent during summer; this species drives silicon cycling in the lake (2). In contrast, weekly data collected in Narragansett Bay in Rhode Island since 1959 reveal that the phytoplankton react with wide fluctuations in composition and timing of the annual biomass

<sup>1</sup>Alfred Wegener Institute for Polar and Marine Research of the Helmholtz Foundation, 27570 Bremerhaven, and the University of Bremen, FB 2, 28334 Bremen, Germany. E-mail: Victor.Smetacek@awi.de <sup>2</sup>U.S. Geological Survey, 345 Middlefield Road, Menlo Park, CA 94025, USA. E-mail: jecloern@usgs.gov

peak to driving forces ranging from large-scale hydrography to the temperature dependence of zooplankton growth (3).

How does phytoplankton performance in other regions fit along this spectrum from monotonous regularity to the verge of chaos (4)? And what will be, or is, the impact of climate change?

Early studies of annual cycles of phytoplankton (5–7) led to the idea that seasonal changes in biomass and species could be attributed to shifts in nutrient and light availability, with grazing pressure increasing in importance in the aftermath of the spring bloom. This annual cycle and its driving forces came to be accepted as the rule by ecologists and modelers, although deviations were reported (8). However, this view assumes that phytoplankton can grow faster than their consumers—zooplankton, pathogens

and dinoflagellate blooms in an era of reduced nutrient inputs. These changes were caused by a shift in the northeast Pacific Ocean from its warm phase to its cold phase, which led to massive immigrations of flatfish and crustaceans into San Francisco Bay, where they reduced clam abundance and weakened their grazing control of phytoplankton growth (9).

Another interesting case is exhibited by data collected off the island of Heligoland in the North Sea since 1967. Here, a 1.5°C rise in winter water temperatures over the past decades resulted in a delay of the spring bloom by some weeks. The shift is attrib-

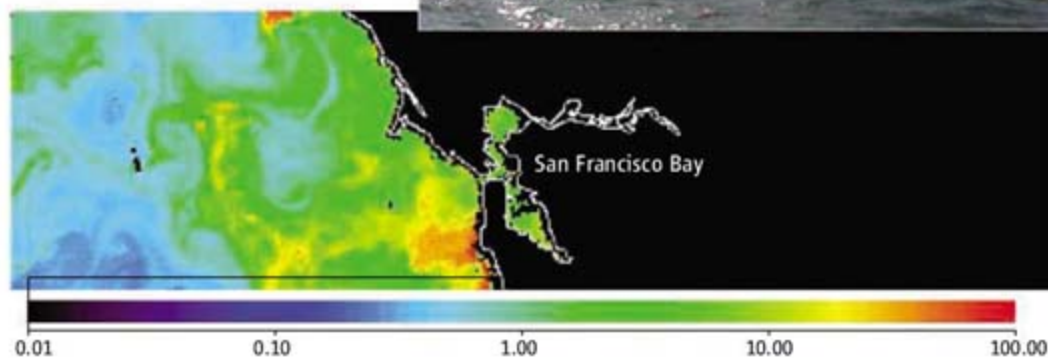
Thus, physicochemical environmental factors (such as temperature, light, and nutrients) set the upper limits to biomass build-up but do not explain why different phytoplankton groups and species replace each other or why their annual maxima occur when and where they do. The fact that most species proliferate for only a few weeks during specific periods suggests that the timing of their appearance is regulated by internal as well as external factors, implying that their life cycles are selected at longer time scales.

Because most phytoplankton species are difficult to identify routinely under a light microscope, it has widely been assumed that many species are opportunistic and respond to favorable growth conditions at any season. However, detailed investigations of the “same” species collected from different periods of the year in the Bay of Naples, Italy, using molecular tools and backed up by detailed visual examination and mating experiments indicate that these comprise distinct, but cryptic, species (12, 13).

The results suggest that species-specific life cycle properties have evolved in individual phytoplankton species in a manner analogous to that in land plants (14). However, it is unclear how environmental fluctuations select phytoplankton species by regulating processes such as sexuality, formation and germination of resting stages, aggregation and sinking, and deployment of defense mechanisms that deter predators and competitors and ward off infection.

The poleward retreat of phytoplankton species with a minimum temperature requirement has been reported, as has the poleward spread of warm-water species. However, most species are temperature-tolerant, and wholesale poleward encroachment of entire coastal ecosystems is unlikely. The response of phytoplankton to ongoing climate change also depends on the geomorphology and hydrology of the respective site. Thus, seasonal shifts in snow-melt and rainfall in the catchment areas change patterns of river discharge, which in turn affect flushing and nutrient delivery rates (15) but also disrupt life cycles of dominant species geared to the former flushing regime. Basin-scale oscillations in the adjoining oceanic regime, modified by global warming, can have similarly drastic effects on coastal regions (9).

In addition to these effects, the structure of coastal food webs down to the level of primary producers will have changed as a result of severe depletion of commercial fish and



**Estuaries and oceans as coupled systems.** Images taken with the MODIS satellite capture episodes of high chlorophyll a concentration in the coastal Pacific Ocean (bottom). Marine plankton and fish are transported into San Francisco Bay (top) through the Golden Gate. Climate-driven variability in the annual recruitment and immigration of predators induces large variability in seasonal algal blooms in San Francisco Bay (9). Similar shifts in seasonality, species, and phytoplankton biomass are common in coastal waters, but the forces driving these shifts are not well understood.

(viruses), and parasites. This is not supported by physiological considerations. Many unicellular phytoplankton consumers divide faster than their prey in the presence of an adequate food supply. Clearly, the mechanisms keeping consumers and pathogens at bay also need to be considered.

The example of San Francisco Bay (see the figure) illustrates how shifting grazing pressure can radically change phytoplankton annual cycles. For two decades, a spring diatom bloom was followed each year by low phytoplankton biomass in the presence of high nutrient concentrations, a result of strong top-down control by clams and other bottom-dwelling suspension feeders during summer and autumn. This annual cycle changed abruptly in 1999 with the appearance of an autumn bloom, a new pattern that has persisted and led to increases in phytoplankton biomass

uted to grazing pressure exerted by larger winter and spring zooplankton populations, partly due to an earlier appearance of the plankton-feeding larvae of benthic invertebrates (10). Similar shifts have been reported from Narragansett Bay. These observations imply that temperature rather than the food supply controls grazing pressure in the spring.

In some coastal ecosystems, nutrient (nitrogen and/or phosphorus) concentrations have declined substantially over the past one or two decades, indicating clear responses to improved wastewater treatment and practices to reduce agricultural runoff. However, in contrast to lakes and the Black Sea (11), coherent phytoplankton responses were not apparent, either as synchronous reductions in biomass or shifts to communities characteristic of low-nutrient habitats.



shellfish stocks (16). Drastic changes in community structure due to removal of top predators or herbivores have been shown in many terrestrial, lake, and marine benthic ecosystems, but there are few examples from marine phytoplankton (11, 17). It is unlikely that marine pelagic ecosystems differ fundamentally from all the others. Long-term monitoring strategies will thus have to be broadened to encompass the full range of abiotic and biotic factors that shape the annual cycles of phytoplankton. Detailed interdisciplinary studies of the

complex interactions characteristic of coastal ecosystems need to be undertaken before global changes obliterate the remnants of the baseline.

#### References

1. American Geophysical Union Chapman Conference, *Long-Time Series Observations in Coastal Ecosystems: Comparative Analyses of Phytoplankton Dynamics on Regional to Global Scales*, Rovinj (Croatia), 8 to 12 October 2007.
2. S. C. Maberly *et al.* *Freshwater Biology* **31**, 19 (1994).
3. T. J. Smayda, *ICES J. Mar. Science* **55**, 562 (1998).
4. E. Benincà *et al.*, *Nature* **451**, 822 (2008).
5. V. Hensen, *Komm. Wiss. Unters. Deutschen Meere* **V**, 12-16, 1 (1887).
6. E. L. Mills, *Biological Oceanography: An Early History*

- 1860-1970 (Cornell Univ. Press, Ithaca, NY, 1989).
7. V. Smetacek, *Estuaries* **8**, 145 (1985).
8. G. A. Riley, *Deep-Sea Res.* **3** (suppl.), 224 (1955).
9. J. E. Cloern *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 18561 (2007).
10. K. Wiltshire, talk presented at the Chapman Conference on *Long Time-Series Observations in Coastal Ecosystems*.
11. A. E. Kideys, *Science* **297**, 1482 (2002).
12. A. Amato *et al.*, *Protist* **158**, 193 (2007).
13. D. Sarno *et al.*, *J. Phycol.* **41**, 151 (2005).
14. R. A. Cheke, *Science* **318**, 577 (2007).
15. I. T. Stewart *et al.*, *J. Climate* **18**, 1136 (2005).
16. R. A. Myers, B. Worms, *Nature* **423**, 280 (2003).
17. M. L. Pace *et al.*, *TREE* **14**, 483 (1998).

10.1126/science.1151330

## SIGNAL TRANSDUCTION

# Sweet Conundrum

Morris J. Birnbaum

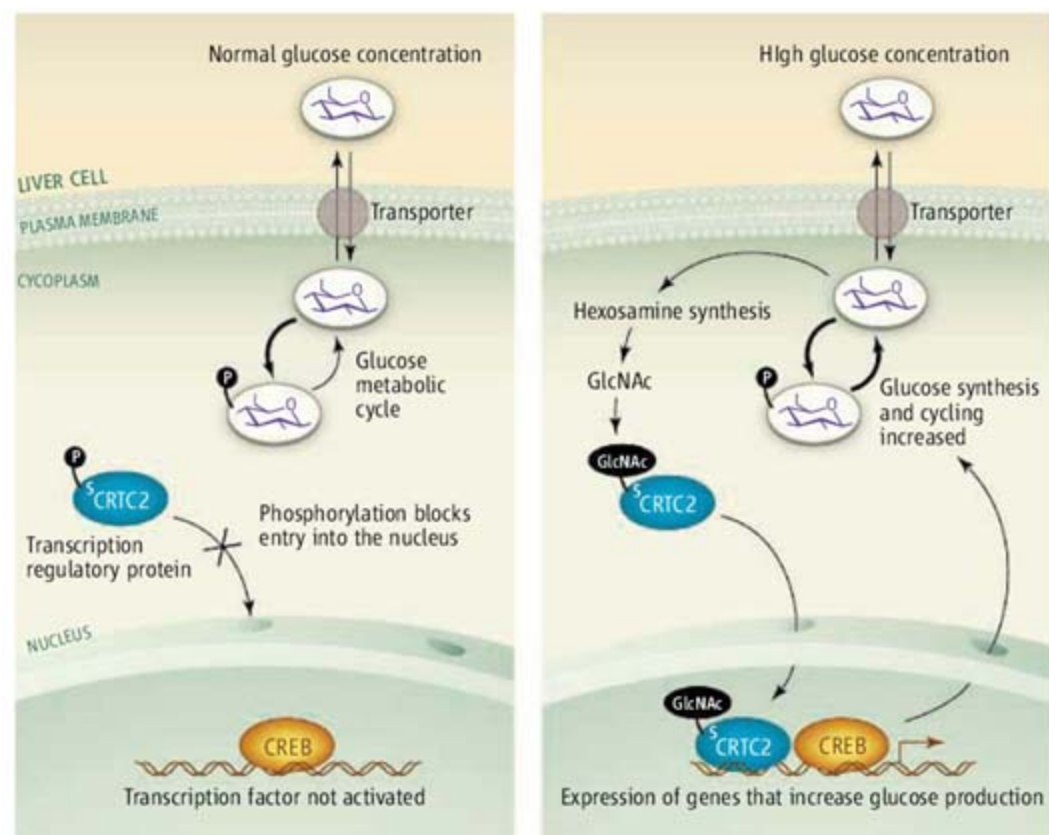
Simple substrates such as glucose and amino acids not only supply energy for growth and biological functions, they also convey information, alerting the organism to the nutritional state of the cellular environment. Thus, glucose not only provides potential energy for generating adenosine 5'-triphosphate in unicellular organisms, but also instructs them to repress the synthesis of enzymes that metabolize alternative substrates. In animals, when the concentration of glucose in the blood rises, the pancreas secretes insulin, the hormone that shuts down glucose production in the liver and stimulates uptake and storage of the sugar in the liver, adipose tissue, and muscle. But when the pancreas can no longer keep pace with increased demand for insulin, hyperglycemia results, a hallmark of diabetes mellitus. The persistently elevated blood sugar concentration eventually provokes the adverse consequences associated with diabetes, through mechanisms still imperfectly understood, but clearly involving an interplay of glucose acting both as a substrate and signaling molecule in cells. On page 1402 in this issue, Dentin *et al.* (1) describe a previously unknown cell signaling pathway by which increased concentration of circulating glucose accelerates its own production by the liver. This seemingly contrary cellular response in the liver to circulating glucose provides further evidence that we must still tease out the complex details of this sugar's metabolism before we can understand how it might be managed under disease conditions.

A cardinal feature of type 2 diabetes mellitus (formerly called adult-onset diabetes

mellitus) is its self-amplifying nature. As increased blood glucose and lipid concentrations worsen, they directly exacerbate the disease. In clinical studies, aggressive insulin treatment of diabetics reduced blood glucose concentrations, interrupting the cumulative causation and lessening the need for treatment (2). Mice rendered diabetic through a mutation that blocks glucose uptake by muscle display a secondary defect—the inability

High blood glucose concentration causes a transcription factor to be modified with a sugar moiety, triggering the liver to produce yet more glucose.

of insulin to suppress glucose production in the liver. In these animals, poisoning the glucose reabsorption process in the kidneys reduces the blood glucose concentration while leaving the underlying defect unchanged and notably ameliorates liver insulin resistance (3). These studies argue that hyperglycemia itself brings about insulin resistance in liver, one consequence of "glucose toxicity." Indeed, although the detrimental effects



**Serine switch.** Phosphorylation (p) on serine(s) residue 171 of the transcriptional regulatory protein CRTC2 blocks its effect on glucose production. However, when blood glucose concentration increases, serine 171 becomes glycosylated, allowing CRTC2 to enter the nucleus, activate the transcription factor CREB, and induce the expression of proteins that increase glucose production. GlcNAc, N-acetylglucosamine.

Institute of Diabetes, Obesity and Metabolism, University of Pennsylvania, Philadelphia, PA 19104, USA. E-mail: birnbaum@mail.med.upenn.edu

of glucose on insulin responsiveness in muscle and on insulin secretion from pancreas are well established, the study by Dentin *et al.* is one of the few to examine hyperglycemic effects in liver. Nonetheless, the existence of glucose toxicity raises the intriguing question of whether the cellular signaling pathways responsible for the adverse effects of glucose are the same as those involved in normal glucose signaling.

As extracellular glucose concentration varies, intracellular signaling pathways adjust accordingly. The glucose signaling pathway studied by Dentin *et al.* was originally described as the hexosamine biosynthetic pathway in the liver. This pathway generates uridine diphosphate (UDP)-*N*-acetylglucosamine, a crucial intermediate in the synthesis of the carbohydrate moiety of complex glycoproteins (4). In 1991, Marshall and colleagues proposed that this same pathway might respond to persistently elevated extracellular glucose concentration by causing insulin resistance in fat cells, though at that time the mechanism was unknown (5). This "hexosamine hypothesis" for the generation of insulin resistance has since been applied to other organs as well as to dysfunctional insulin secretion, and the current favored mechanism is that UDP-*N*-acetylglucosamine drives the enzymatic, O-linked (modification of a hydroxyl group) glycosylation of serine and threonine residues in intracellular proteins (4, 5). This posttranslational modification is widely distributed in normal cells, particularly in proteins of the nuclear pore and those associated with chromatin. Dentin *et al.* add to the list of such modified proteins the transcriptional regulatory protein called transducer of regulated cAMP response element-binding protein 2 (TORC2, also known as CRTC2). O-linked glycosylation of CRTC2 activates glucose production in the liver and is an example of how this modification of proteins serves a clear signaling function in controlling metabolism in the liver.

In the nucleus, CRTC2 associates with CREB (cyclic AMP response element-binding protein), a transcription factor that activates the expression of genes that control glycolysis and gluconeogenesis (6). CRTC2 enhances the transcription of genes encoding proteins critical to gluconeogenesis, in particular the enzyme glucose-6-phosphatase. This enzyme catalyzes the terminal step in glucose production by the liver and is normally negatively regulated by insulin—when insulin concentration is high, total phosphatase concentration is low (7). The cellular energy sensor adenosine 5'-monophosphate (AMP)-activated protein kinase also reduces

the concentration of glucose-6-phosphatase, but it phosphorylates CRTC2 on serine 171, inducing its translocation to the cytoplasm where it can no longer affect transcription (8). Because *N*-acetylglucosamine attaches itself to the same serine that is phosphorylated, glycosylation serves as a switch to prevent the inhibition of CRTC2. The ultimate result is that abnormally high blood sugar concentration activates enzymes designed to release more glucose (7) (see the figure).

So, is regulation of CRTC2 by O-linked glycosylation a normal process that becomes maladaptive during states of nutritional excess, or is it wholly pathological, where pure coincidence allows the hexosamine pathway to block phosphorylation? Both alternatives seem unlikely. It is difficult to see why glucose would induce its own synthesis, and it seems improbable that such an elegant control mechanism would evolve by chance without some selective pressure. Perhaps the answer lies in the complexities of liver metabolism, in which viewing glucose consumption and production as two opposing, linear pathways is too simplistic. For example, glucose-6-phosphatase not only serves in the net production of glucose,

but also catalyzes the cycle that converts glucose back to glucose-6-phosphate (and back again). Indeed, under at least some conditions, increases in extracellular glucose concentration have been associated with an increase in this cycling rather than a change in net glucose-6-phosphatase flux (9). Thus, even important new insights such as those provided by Dentin *et al.* will require more study before we can understand how they integrate into the complex metabolism of an intact organism under normal and pathological conditions.

#### References

1. R. Dentin, S. Hedrick, J. Xie, J. Yates III, M. Montminy, *Science* **319**, 1402 (2008).
2. W. T. Garvey, J. M. Olefsky, J. Griffin, R. F. Hamman, O. G. Kolterman, *Diabetes* **34**, 222 (1985).
3. J. K. Kim *et al.*, *J. Clin. Invest.* **108**, 153 (2001).
4. C. Slawson, M. P. Housley, G. W. Hart, *J. Cell Biochem.* **97**, 71 (2006).
5. S. Marshall, in *Sci. STKE* **2006**, re7 (2006).
6. A. Cheng, A. R. Saltiel, *Bioessays* **28**, 231 (2006).
7. D. Massillon, N. Barzilai, W. Chen, M. Hu, L. Rossetti, *J. Biol. Chem.* **271**, 9871 (1996).
8. D. G. Hardie, S. A. Hawley, J. W. Scott, *J. Physiol. (London)* **574**, 7 (2006).
9. S. Efendic, A. Wajngot, M. Vranic, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2965 (1985).

10.1126/science.1155915

#### COMPUTER SCIENCE

## Science 2.0

Ben Shneiderman

Traditional scientific methods need to be expanded to deal with complex issues that arise as social systems meet technological innovation.

The growth of the World Wide Web and the spread of cell phones and WiFi continues to reorder whole disciplines and industries. Entrepreneurs, policy-makers, and researchers have recognized that increased collaboration through these socio-technical systems offers compelling opportunities for business, education, national security, and beyond (1). It is time for researchers in science to take network collaboration to the next phase and reap the potential intellectual and societal payoffs (2–4).

Successful scientific collaborations among genomic researchers, engineering innovations through open-source software, and community-based participation in cultural heritage projects are all early indicators of the transformative nature of collaboration (5).

Department of Computer Science, University of Maryland, College Park, MD 20742, USA. E-mail: [ben@cs.umd.edu](mailto:ben@cs.umd.edu)

eBay, Amazon, and Netflix have already reshaped consumer markets, while political participation and citizen journalism are beginning to change civil society. Patient-centered medical information and secure electronic health records are improving health care while creating opportunities for clinical research. MySpace and Facebook encourage casual social networks, but they may soon play more serious roles in facilitating emergency/disaster response (6). Social media platforms such as Wikipedia, flickr, and YouTube are also stunning success stories of Web-based contributions.

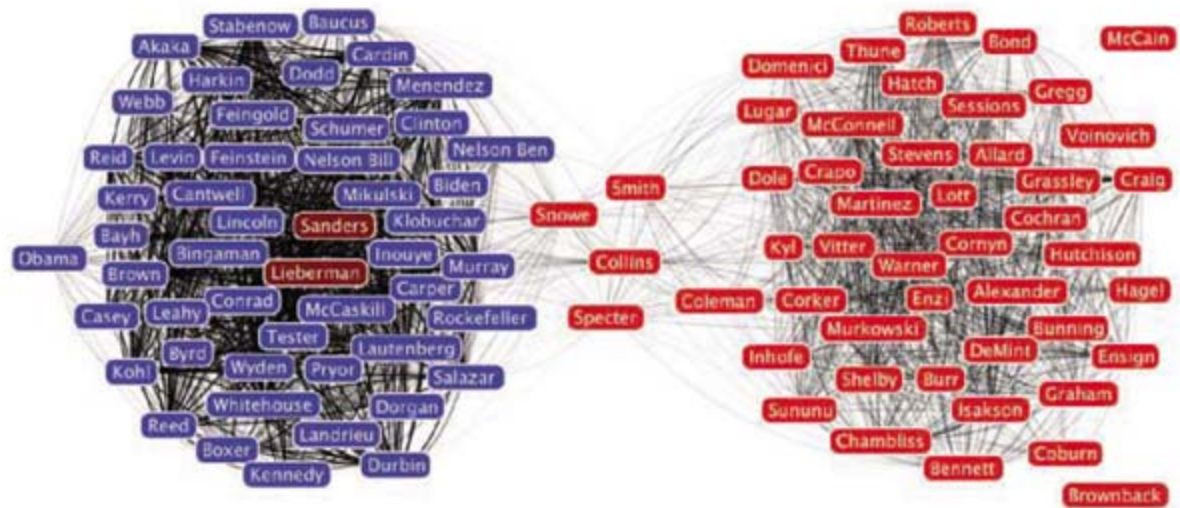
Understanding these collaboration-centered socio-technical systems could accelerate their adoption and raise their benefits. However, researchers will need to develop new ways of studying these complex interactions. Science 1.0 will continue to be important, but new kinds of science, which I call

Science 2.0, are needed to study the integrated interdisciplinary problems at the heart of socio-technical systems. Science 2.0 will be especially important to meet the design challenges in secure voting, global environmental protection, energy sustainability, and international development among many others.

The guiding strategies of Science 1.0 are still needed for Science 2.0: hypothesis testing, predictive models, and the need for validity, replicability, and generalizability. However, the Science 2.0 challenges cannot be studied adequately in laboratory conditions because controlled experiments do not capture the rich context of Web 2.0 collaboration, where the interaction among variables undermines the validity of reductionist methods (7). Moreover, in Science 2.0 the mix of people and technology means that data must be collected in real settings (see the figure). Amazon and Netflix became commercial successes in part because of their frequent evaluations of incremental changes to their Web site design as they monitored user activity and purchases.

Researchers who wish to foster online health care information groups or citizen journalism, for example, need fresh research methods and theories (8, 9). Individual outcomes are difficult enough to study, but understanding why the Google, YouTube, or Facebook communities succeeded in the face of lively competition is still more challenging. These socio-technical systems are best studied at scale, in the real world, by rigorous observations (studying successes and failures), carefully chosen interventions (changing interfaces or privacy rules), and ambitious data collection (analyzing all public user activity). When adequately replicated, these quantitative and qualitative empirical studies can lead to predictive models and effective simulations that guide future designers and researchers.

Science 1.0 heroes such as Galileo, Newton, and Einstein produced key equations that describe the relationships among gravity, electricity, magnetism, and light. By contrast, Science 2.0 leaders are studying trust, empathy, responsibility, and privacy. The great adventure for the next 400 years will be to define, measure, and predict the interaction among these variables so as to accelerate scientific discovery, engineering innovation, e-commerce, and education (10). The fivefold growth of research on privacy and trust is apparent in the past 5 years, whereas empathy



**Political networking.** Collaboration between pairs of U.S. senators is shown by connecting links. The Democratic senators (blue) are at the left and Republican senators (red) at the right; Sanders and Lieberman (magenta) are independents. Brownback and McCain were campaigning for the presidency and did not vote often enough to be linked. Science 2.0 must develop tools like this to analyze human relationships and collaborations (16).

and responsibility are just beginning to capture attention (11).

Science 2.0 researchers are adopting observational and case study methods as they collect quantitative and qualitative data to gain support for their hypotheses about whether trust increases empathy and whether privacy promotes responsibility (12, 13). Their work methods are in harmony with research initiatives on Web science (14), creativity support tools, online education (15), and socially networked communities, among others.

Advancing Science 2.0 will require a shift in priorities to promote integrative thinking that combines computer science know-how with social science sensitivity. Science 2.0 researchers who develop innovative theories, hypothesis testing based on case study research methods, and new predictive models are likely to lead the way. The quest for empirical validity will drive research beyond what laboratory-based controlled studies can provide, while replicability and generalizability will be achieved with greater effort through multiple case studies. Just as technology-centered researchers measured progress in petabytes of storage or petaflops of processing power, collaboration-centered researchers will measure the growth of peta-collabs of cooperation and peta-contribs of assistance.

Science 1.0 remains vital, but this ambitious vision of Science 2.0 will affect research funding, educational practices, and evaluation of research outcomes. Science funding agencies will face resistance as they promote a transformation that seeks to make a safe space for Science 2.0. Scientific journal editorial boards and conference program committees are already shifting their attention to new topics and opening their doors to new scientific research methods. Pioneering edu-

cators have begun changing their curricula, focusing on collaboration strategies and teaching new research methods. The innovators are courageously taking on new challenges, but they should be ready for the resistance to novel ideas that has always been part of science. In that way, Science 2.0 is part of a great tradition.

#### References and Notes

1. E. Mumford, *Requir. Eng.* **2**, 59 (2000).
2. R. Hiltz, M. Turoff, *The Network Nation* (MIT Press, Cambridge, MA, 1993).
3. B. Shneiderman, *Leonardo's Laptop: Human Needs and the New Computing Technologies* (MIT Press, Cambridge, MA, 2002).
4. W. S. Bainbridge, *Science* **317**, 472 (2007).
5. N. Bos et al., *J. Computer-Mediated Commun.* **12**, 2 (2007).
6. B. Shneiderman, J. Preece, *Science* **315**, 944 (2007).
7. L. A. Suchman, *Plans and Situated Actions: The Problem of Human-Machine Communication* (Cambridge Univ. Press, Cambridge, 1987).
8. D. Maloney-Krichmar, J. Preece, *ACM Trans. Computer-Human Interact.* **12**, 201 (2005).
9. H. T. Welsler, E. Gleave, D. Fisher, D., M. Smith, *J. Soc. Struct.* **8**, 2 (2007).
10. G. Fischer, E. Scharff, Y. Ye, in *Social Capital and Information Technology*, M. Huysman, V. Wulf, Eds. (MIT Press, Cambridge, MA, 2004), pp. 355–399.
11. Based on a search of abstracts in the ACM Digital Library: <http://portal.acm.org/dl.cfm>.
12. R. K. Yin, *Case Study Research: Design and Methods* (Sage, Thousand Oaks, CA, ed. 3, 2003).
13. V. Kaptelinin, B. Nardi, *Acting with Technology: Activity Theory and Interaction Design* (MIT Press, Cambridge, MA, 2006).
14. T. Berners-Lee, W. Hall, J. Hendler, N. Shadbolt, D. J. Weitzner, *Science* **313**, 769 (2006).
15. S. R. Hiltz, R. Goldman, Eds., *Learning Online Together: Research on Asynchronous Learning Networks* (Erlbaum, Mahwah, NJ, 2005).
16. More information about the network diagram including a movie version is available at [www.cs.umd.edu/hcil/socialaction](http://www.cs.umd.edu/hcil/socialaction).
17. I thank J. Preece, R. Rice, R. Hiltz, M. Smith, G. Fischer, J. Golbeck, and K. Borner for comments.

## RETROSPECTIVE

## Joshua Lederberg (1925–2008)

Stephen S. Morse

Joshua Lederberg was one of the great scientists of our age. With his death on 2 February, the world has lost one of its foremost scientific intellects, as well as an extraordinary humanitarian.

These qualities were apparent early in his life. Descended from a long line of rabbis, Josh was expected to continue the tradition. But at age 10, when he decided to become a scientist instead, his father reassured him that “All those who seek the truth are doing God’s work.” And when two high-school classmates died in World War II, he vowed to make up for their sacrifice by his own life and work (he was too young to be in the war himself).

For the many people he helped over the years, Josh was the ultimate mentor. Twenty years ago at a Rockefeller University gathering that I attended, his wife Marguerite reminded him that he had some virology questions. So began a continuing conversation. Following up, I sent Josh a memo about “emerging viruses,” and shortly afterwards received a handwritten note declaring, “The problem of emerging viruses is one that must be addressed at the highest levels.” This led to the “Conference on Emerging Viruses” that John LaMontagne and I organized at the National Institutes of Health in 1989, in which Josh actively participated. He organized and chaired a landmark Institute of Medicine study on “Emerging Infections: Microbial Threats to Health in the United States” (published in 1992), and the subsequent “Forum on Emerging Infections” (now the “Forum on Microbial Threats”), and invited me to work with him on these committees.

Born in Montclair, New Jersey, in 1925, Josh grew up in New York City. After graduating from Columbia University in 1944, and 2 years in medical school, he decided on a career in research. Transferring to graduate school at Yale University, he studied with Edward Tatum and received his Ph.D. in 1947. His receipt of the Nobel Prize in 1958 was based largely on his graduate research on the organization of the genetic material in bacteria and on genetic recombination. His work on genetic transfer by conjugation opened the door to molecular biology with bacteria. At the University of Wisconsin, he and his grad-

uate student Norton Zinder discovered transduction, a mechanism of genetic exchange in bacteria facilitated by bacteriophage. Many bacterial virulence factors are now known to be acquired in this way. In 1959, Josh moved west to start the Department of Genetics at Stanford University. He returned to New York in 1978 as the fifth president of The Rockefeller University. When he officially retired as president in 1990, he reopened his lab, and nurtured many creative young scientists. It was not uncommon to spot him on campus wearing a short-sleeved shirt and a Rockefeller baseball cap.

His abiding biological interest was in the sources of genetic variation and the effects of natural selection. Fascinated by the complex relationships between humans and microbes, he remained concerned that future plagues might devastate humankind.

Josh also had a gift for the apt metaphor. He noted that “the historiography of epidemic disease is one of the last refuges of the concept of special creationism.” In recent years, he adopted the metaphor of the food chain, comparing humans and microbes and questioning which was really at the top of the chain. He referred to a “World Wide Web of microbial genomes,” prefiguring metagenomics. After hearing him speak, the listener might be struck later by profound observations that had initially gone unnoticed.

A great visionary, Josh was endlessly curious and creative. His footprints can be found in many fields. In addition to his wide interests in biology, he had long been interested in the use of computers in science and was a pioneer in the field of bioinformatics. At Stanford, he collaborated on DENDRAL, an early project in chemical-analysis expert systems using computers. Later, he helped develop the concept of the “collaboratory,” a virtual collaboration across geographic boundaries. The advent of the Internet is now helping to bring that dream to fruition. Appropriately, electronic copies of many of

The interests of a visionary scientist ranged from genetic details of the microbial world to new frontiers in computer science and space exploration.



his personal papers are available on the Web at the National Library of Medicine.

Perhaps less known is his interest in exobiology (a term he coined). After the launch of the *Sputnik* satellite in 1957, Josh wanted to ensure that biological science was part of space research programs. His earliest scientific efforts with the National Aeronautics and Space Administration were directed toward developing approaches to identify life on other planets. Later, he advised the agency on precautions for

decontaminating returning spaceships.

Josh was a central figure in science policy, and was generous with his time whenever the government or colleagues needed his advice. Before back pain and complications of orthopedic surgery precluded travel, he would regularly shuttle between New York and Washington, DC, sometimes making several round trips a week. At one meeting, he characteristically joked, “We always meet in the most expected places.”

In the last few years, despite often being in pain, his eyes lit up whenever the conversation turned to science. That Josh’s contributions to science and science policy were of the highest order is indicated by his many awards and honors. In addition to the Nobel Prize, these included the National Medal of Science, election to the National Academy of Sciences, charter membership of the Institute of Medicine, and the Presidential Medal of Freedom. But these only begin to hint at his broad interests and contributions. He was a longtime board member of the Council on Foreign Relations, held the Order of Arts and Letters from France, and was an avid reader, from scientific literature to military history of World War II and the *Times Literary Supplement*.

He leaves behind Marguerite, their two children and grandchildren, and devoted colleagues. But he also leaves generations of scientists whom he encouraged and prodded to new heights. With his brilliance and generosity of spirit, he brought light everywhere he went.

10.1126/science.1156612

Mailman School of Public Health, Columbia University, New York, NY 10032, USA. E-mail: [ssm20@columbia.edu](mailto:ssm20@columbia.edu)

# An Oncogene-Induced DNA Damage Model for Cancer Development

Thanos D. Halazonetis,<sup>1\*</sup> Vassilis G. Gorgoulis,<sup>2</sup> Jiri Bartek<sup>3</sup>

Of all types of DNA damage, DNA double-strand breaks (DSBs) pose the greatest challenge to cells. One might have, therefore, anticipated that a sizable number of DNA DSBs would be incompatible with cell proliferation. Yet recent experimental findings suggest that, in both precancerous lesions and cancers, activated oncogenes induce stalling and collapse of DNA replication forks, which in turn leads to formation of DNA DSBs. This continuous formation of DNA DSBs may contribute to the genomic instability that characterizes the vast majority of human cancers. In addition, in precancerous lesions, these DNA DSBs activate *p53*, which, by inducing apoptosis or senescence, raises a barrier to tumor progression. Breach of this barrier by various mechanisms, most notably by *p53* mutations, that impair the DNA damage response pathway allows cancers to develop. Thus, oncogene-induced DNA damage may explain two key features of cancer: genomic instability and the high frequency of *p53* mutations.

In medicine, neoplasia (a Greek word meaning “new growth”) refers to new tissue composed of cells with the heritable capacity to grow beyond their normal confines. Neoplasias can be benign or malignant. The former remain localized, whereas the latter, also referred to as cancer, invade the host tissues. Some benign neoplasias are precancerous lesions and develop over time into overt cancers. Important goals of cancer research in the molecular biology era have been to understand the genetic basis of cancer and to explain its progressive nature. Here, we review a model that may contribute to better understanding of two features shared by most cancers: genomic instability and *p53* mutations.

## DNA DSBs in Cells Expressing Activated Oncogenes and in Human Precancerous Lesions and Cancers

Until recently, identifying DNA double-strand breaks (DSBs) with high sensitivity was not trivial. However, the identification of proteins that become recruited to large chromatin domains flanking the DNA DSBs has made it possible to visualize these breaks indirectly by immunofluorescence (1). (Additional references are listed in the Supporting Online Material, grouped by the section in this review to which they pertain.) The intracellular localization of one such protein, *p53* binding protein 1 (53BP1), was examined in cancer cell lines grown in tissue culture and revealed 53BP1 nuclear foci (about 10 to 20 per cell), whose presence is indicative of DNA

DSBs. No such foci were present in proliferating normal cells. Because cancer cells are typically not defective in DNA DSB repair, these results suggested a continuous cycle of DNA DSB formation and repair (1, 2). Further, in the panel of cancer cell lines examined, those cell lines with the highest number of 53BP1 foci per cell had *p53* mutations, possibly reflecting a selection for *p53* inactivation, because *p53* induces apoptosis or senescence in response to DNA DSBs (1, 2).

The analysis of cancer cells in tissue culture is informative but may not accurately represent the state of cancer cells in human patients. Are DNA DSBs continuously being generated in human cancers? If so, at what stages of cancer development? And does DNA DSB formation precede the loss of *p53* function, as would be expected if the DNA DSBs select for *p53* inactivation? To address these questions, several research groups analyzed precancerous and cancerous lesions from human patients. In the precancerous lesions, before *p53* mutations were acquired, 53BP1 localized at foci and histone H2AX, ataxia telangiectasia (ATM), Chk2, and *p53* were phosphorylated, suggesting the presence of DNA DSBs (3, 4). In cancers, evidence for the presence of DNA DSBs, such as phosphorylated histone H2AX and 53BP1 foci, was again present, but the DNA damage checkpoint pathway was compromised, most often by *p53* mutations and less often by loss of expression of various checkpoint proteins, such as ATM, 53BP1, Chk2, or *p53* (2–4).

The presence of DNA damage was a feature that could distinguish precancerous lesions and cancers from normal tissues, irrespective of their proliferation rate (2–4). What could be causing the induction of DNA DSBs in the precancerous lesions and cancers? Eroded telomeres or mutations targeting genes required for genome integrity could be responsible, but neither of these characterizes all precancerous lesions (5–8).

Instead, the presence of activated oncogenes distinguishes precancerous lesions from normal tissues, and in various model systems activated oncogenes are capable of inducing both DNA DSBs and a DNA damage response (DDR). For example, activated *ras* induces DNA DSBs in NIH3T3 fibroblasts within a single cell cycle, and other oncogenes, including *myc*, *cyclin E*, *mos*, *cdc25A*, and *E2F1*, have similar effects in various cell types and in animal models (4, 9–11). In a human skin xenograft hyperplasia model, sustained delivery of growth factors also led to induction of DNA DSBs and a DDR, in the absence of telomere erosion (3). In all these models, the majority of cells overexpressing the oncogene exhibited a DDR, suggesting that a second stochastic event was not required for induction of DNA damage.

## A Tumorigenesis Barrier in Human Precancerous Lesions

In various cell and animal models, activation of the DNA damage checkpoint induces *p53*-dependent cell cycle arrest, apoptosis, or senescence (12). The question then arises whether apoptosis and senescence are present in human precancerous lesions and, if so, whether they subside during cancer progression, when defects in the *p53* pathway are acquired. Indeed, the precancerous tissues show evidence of apoptosis or senescence, and both of these are suppressed during cancer progression (3, 4, 10, 11, 13).

This general pattern is observed in multiple tumor types, but with interesting tissue-specific differences (3, 10, 13). In the lung, the normal bronchial epithelium has a very low proliferation index and negligible apoptosis; the precancerous lesions exhibit high levels of apoptosis and a high proliferation index; whereas lung carcinomas have a high proliferation index but low levels of apoptosis. In the skin, normal melanocytes exhibit a low proliferation index and negligible senescence; dysplastic nevi, the precancerous lesions, exhibit senescence and a low proliferation index (because senescent cells do not proliferate); and in melanoma there is loss of the senescent phenotype and a high proliferation index. In colon, the normal tissue has a very high proliferation index; progression to adenoma, the precancerous lesion, leads to senescence and a significant decrease in the proliferation index; whereas further progression to carcinoma is associated with escape from senescence and a high proliferation index similar to that of the normal tissue. Plotting the apoptosis or senescence indices of these three tissues (3, 10, 13) shows a sharp peak in precancerous lesions, representing a barrier that is raised specifically in that stage of cancer development and which is eroded upon further tumor progression (Fig. 1).

The concept of a tumorigenesis barrier acting in precancerous lesions contrasts with earlier models, which considered that precancerous lesions were less aggressive than cancers simply

<sup>1</sup>Department of Molecular Biology and Department of Biochemistry, University of Geneva, CH-1205 Geneva, Switzerland. <sup>2</sup>Department of Histology and Embryology, School of Medicine, University of Athens, GR-11527 Athens, Greece. <sup>3</sup>Institute of Cancer Biology and Centre for Genotoxic Stress Research, Danish Cancer Society, DK-2100 Copenhagen, Denmark.

\*To whom correspondence should be addressed. E-mail: [Thanos.Halazonetis@molbio.unige.ch](mailto:Thanos.Halazonetis@molbio.unige.ch)

because they had fewer oncogenic mutations (14). Instead, the oncogene-induced DNA damage model suggests that the less aggressive nature of precancerous lesions is in part due to the tumorigenesis barrier imposed by the DNA damage checkpoint. This concept is best illustrated by the example of colon adenomas, which have a lower proliferation index than that of normal colon (Fig. 1).

### The DNA Damage Checkpoint as an Important Mediator of the Tumorigenesis Barrier

Much evidence points to p53 as a key protein imposing the tumorigenesis barrier in precancerous lesions: Apoptosis and senescence are known p53-dependent responses, escape from the tumorigenesis barrier correlates with p53 mutations in human cancers, and the transition from precancerous lesions to cancer is accelerated in p53-deficient mice (3, 4, 12, 13). In fact, in some p53-null mouse models, tumor development bypasses the senescence stage (13). The question that remains to be answered is whether, in precancerous lesions, p53 is activated by the DNA damage checkpoint pathway, as the oncogene-induced DNA damage model predicts, or by one of the other pathways that activate p53, with the most serious contender being the pathway that includes the alternative reading frame (ARF) tumor suppressor protein (15).

Perhaps the strongest evidence in favor of p53 being activated by the DNA damage checkpoint pathway in human cancers comes from analysis of clinical material. Analysis of colon and bladder precancerous lesions by immunohistochemistry showed that the senescence markers coincide with the DDR markers (10). The spectrum of genetic mutations present in cancer patients also suggests that p53 is principally activated by the DNA damage checkpoint. Activation of p53 by DNA DSBs is mediated principally by the kinases ATM and Chk2 (12). In an analysis of 518 protein kinase genes in 210 human cancers, the *atm* gene ranked third in terms of mutation frequency (16). Further, heterozygous *chk2* germline mutations predispose one to cancer and can occasionally be associated with a Li-Fraumeni syndrome that is indistinguishable from the one caused by p53 germline mutations (17). Lastly, expression of 53BP1 and MDC1, two DDR proteins that act upstream of Chk2 and p53, is suppressed in subsets of melanomas, breast and lung carcinomas (3). Mutations targeting *arf* are also frequent in human cancer, but they are typically deletions also targeting the tumor suppressor genes *INK4a* and *INK4b* (15). Point mutations targeting *arf* without affecting *INK4a* are found with about one-twentieth the frequency of point mutations targeting exclusively *INK4a*.

A role of the DNA damage checkpoint in mediating a tumorigenesis barrier is also supported by analysis of cultured primary cells overexpressing oncogenes. Such cells exhibit apoptosis or senescence or both, but these responses are suppressed when ATM is inhibited (10, 11). In mouse

tumor models, apoptosis induced by oncogenes, such as *myc*, also depends on the DDR pathway, and inhibiting DDR proteins, such as Chk2, ATM, or Tip60, facilitates tumor progression (10, 11). Tumors also develop in mice deficient for *atm* or *chk2* and in mice in which two residues in p53 that are phosphorylated by ATM and Chk2 are substituted with alanines (18).

Nevertheless, there are also observations that seem to contradict the notion that the DNA damage checkpoint is an important barrier to tumorigenesis. First, why is p53 mutated much more frequently in human cancer than any other DNA damage checkpoint gene? One explanation is that single amino acid substitutions targeting the p53 DNA binding domain can easily inactivate the protein (because of its very low melting temperature) and at the same time generate mutants with dominant-negative activity. Further, unlike inactivation of some upstream DDR proteins, such as ATM, p53 inactivation does not compromise the G2/M-phase DNA damage checkpoint or DNA DSB repair, upon which cancer cells are likely to rely to optimize their survival (12).

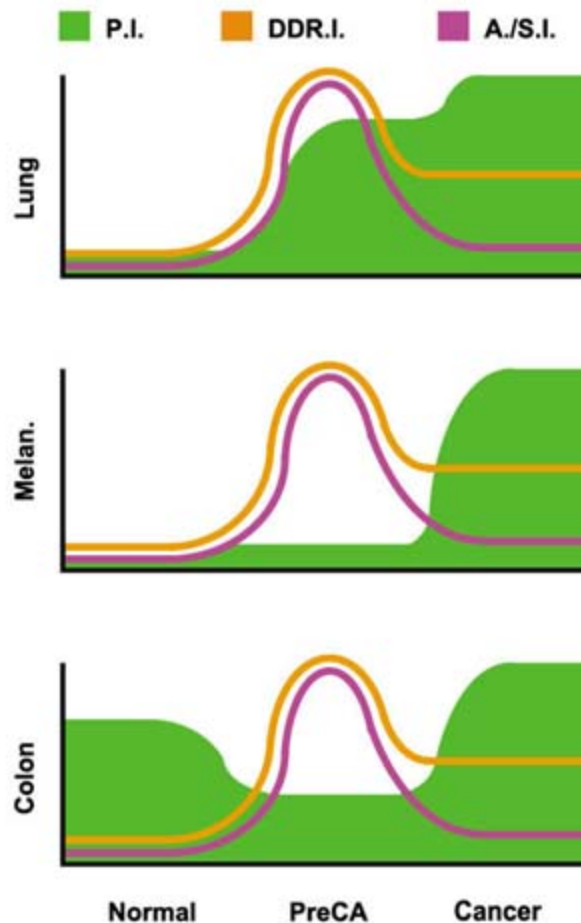
A second question is why knockout of the *atm* and *p53* genes in mice leads to a shorter latency in spontaneous tumor formation (median

survival age of about 55 days) than the individual knockouts (medians of about 120 and 140 days for the *atm* and *p53* knockouts, respectively) (18). Further, why do tumors that develop in *atm*-null, *p53*-heterozygote mice tend to lose the remaining wild-type *p53* allele? These observations appear to contradict the notion that p53 is activated by ATM in precancerous lesions. However, analysis of the median survival ages of mice with all combinations of *atm* and *p53* genotypes shows that the effect of knocking out one or both alleles of *p53* is much smaller in an *atm*-null background than in heterozygote or wild-type *atm* backgrounds, which is consistent with ATM activating the tumor suppressor function of p53 (18). The residual tumor suppressor function of p53 in *atm*-null mice can be explained in part by activation of p53 by ATR, an ATM-related kinase that, like ATM, also responds to DNA damage (12). Further, the residual tumor suppressor function of ATM in *p53*-null mice can be explained by ATM having multiple substrates in addition to p53, some of which, for example, Nbs1 and BRCA1, are also tumor suppressors (12).

A third question relates to the role of *arf* as an alternate pathway activating p53 in cancer.

Indeed, many studies examining the role of *arf* in cultured cells and mouse models support a role in oncogene-induced apoptosis and senescence (15), although there are exceptions. Interestingly, in the same mouse model, a lymphoma induced by a *myc* transgene, lymphomagenesis can be accelerated by knocking out either *arf* or *atm*. This suggests that both the DDR and the *arf* pathways have tumor suppressor functions but leaves open the question whether the tumor suppressor function of *arf* is p53-dependent. Studies to answer this question in mice have led to conflicting conclusions (15).

Some observations suggest that part of the tumor suppressor effect of *arf* is p53-independent. First, in mice, *arf* is a much weaker tumor suppressor than *atm* or *p53*: The median survival ages of mice null for *arf*, *atm*, and *p53* are about 350, 120, and 140 days, respectively (15, 18, 19). This is not consistent with *arf* being the major activator of the tumor suppressor function of p53. Second, analysis of tumor formation in mice with various combinations of *arf* and *p53* genotypes suggests that *arf* and *p53* are acting additively and independently of each other. Specifically, the tumor suppressor function of p53 is not compromised in an *arf*-null background (19), unlike what was observed in an *atm*-null background (18). Third, *arf* may be exerting its tumor suppressor function at a different stage of cancer development than does p53. In a chemical carcinogen-



**Fig. 1.** DNA damage-induced tumorigenesis barrier in human precancerous lesions. The proliferation index (P. I., green), the apoptosis or senescence index (A./S.I., purple), and the DNA damage response index (DDR.I., orange) are shown for normal tissues, precancerous lesions (PreCA), and cancers. For lung tissue, the tumorigenesis barrier is predominated by the apoptotic index; for melanocytes (melan.) and colon, by the senescence index.

esis mouse skin cancer model, inactivation of *arf* promotes formation of large papillomas, whereas inactivation of *p53* promotes progression of papillomas to carcinomas (15).

Contradicting the results presented above (19), two recent studies concluded that *p53* did not have a tumor suppressor function in *arf*-null mice (20). In one study, the function of a tamoxifen-regulatable *p53* protein was turned on or off, and the incidence of radiation-induced lymphoma was monitored. In the other study, 3-methylcholanthrene-induced tumor formation was monitored in mice with two or three wild-type *p53* alleles. In both studies, *p53* status affected tumor kinetics in a wild-type *arf* genetic background but not in an *arf*-null background. However, when tumor kinetics were considered in all genotypes, the results were consistent with *p53* and *arf* acting independently and additively of each other. The status of *p53* had no effect in the *arf*-null background because in these models the tumors develop fast in the absence of *arf*, before any effect of *p53* can become apparent. Interestingly, in the tamoxifen-regulatable *p53* model, administration of tamoxifen for a short time window a few weeks after irradiation, when precancerous cells are presumably present, affords tumor protection. Because the tamoxifen-regulatable *p53* protein requires both tamoxifen and a DNA damage signal to induce apoptosis (20), this result is consistent with the precancerous cells in these mice having DNA DSBs.

**DNA Replication Stress Underlies DNA DSB Formation and Genomic Instability in Human Precancerous Lesions**

A key question that follows from the results reviewed so far relates to the mechanism by which activated oncogenes continuously induce DNA DSBs in human precancerous lesions and cancers. Most oncogenes deregulate entry into the cell cycle and do so by directly or indirectly enhancing the activities of the cyclin-dependent kinases (CDKs) that function in the G1 and S phases of the cell cycle (7, 21). In yeast, deregulation of CDK activity compromises DNA replication and leads to formation of DNA DSBs and genomic instability (22). By analogy, oncogenes could induce a state of DNA replication stress in human precancerous lesions leading to the formation of DNA DSBs (3, 4, 10, 11).

The term DNA replication stress was originally used to describe the state of DNA replication arrest that is induced by

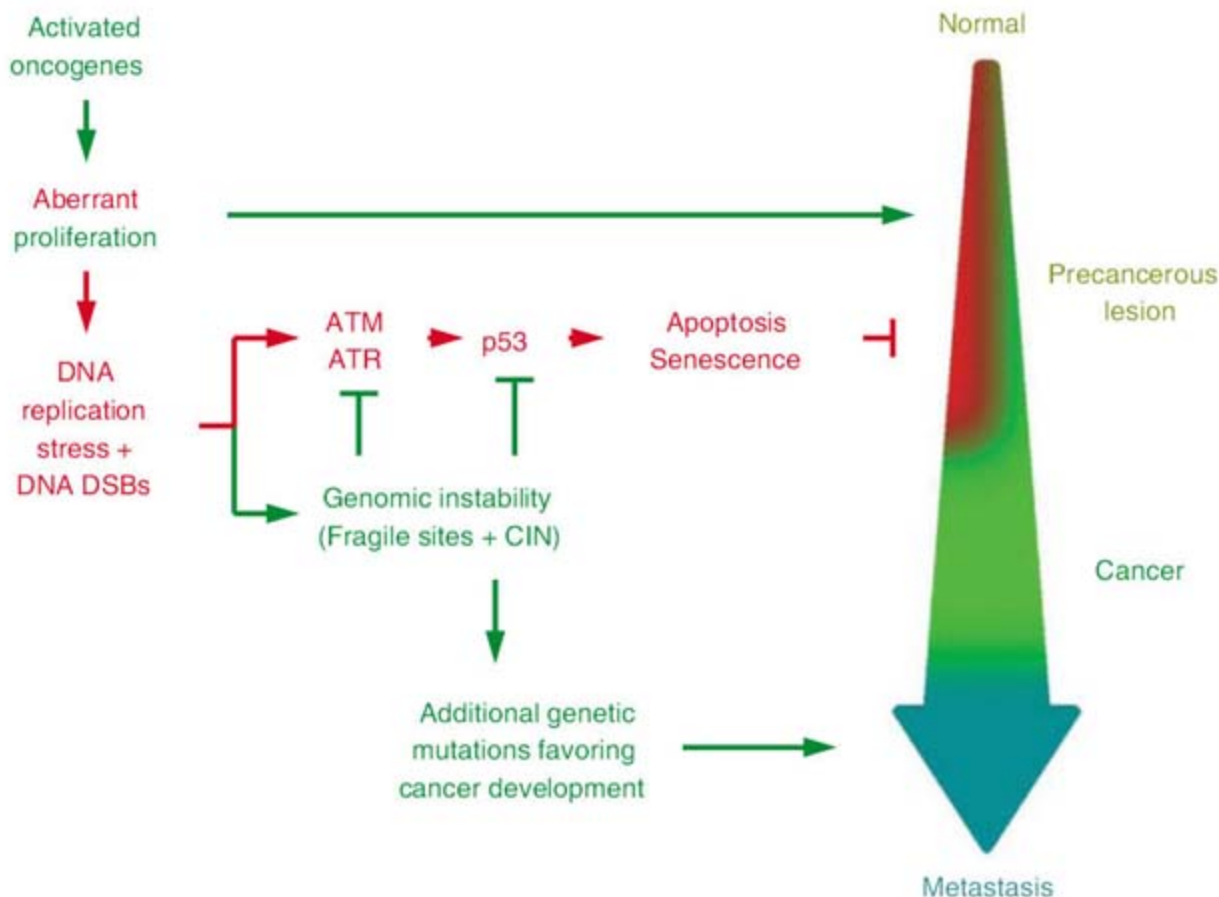
deoxynucleotide depletion and is characterized by activation of the DNA replication checkpoint pathway and inability to complete DNA replication in the absence of this checkpoint. Through a better understanding of the biology of DNA replication forks, it is now possible to define DNA replication stress as any systemic state in the cell that leads to collapse of DNA replication forks, that is, to dissociation of the replication proteins from the DNA. DNA replication stress can be induced by an increase in the number of stalled DNA replication forks (some of which will inevitably collapse) mediated by agents that inhibit DNA replication or by a decrease in the stability of stalled forks (because there is some fork stalling in normal S phases) mediated by deregulation of DNA replication checkpoint proteins. Irrespective of how DNA replication stress is induced, the collapse of DNA replication forks occurs preferentially at specific chromosomal loci called common fragile sites (23). These loci are prone to formation of microdeletions and gross chromosomal rearrangements, because the pathways that are induced to complete DNA replication after fork collapse often involve recombinogenic processes and formation of DNA DSBs.

The presence of DNA replication stress in human precancerous lesions can be monitored by exploiting the fact that the recombinogenic processes that are activated after fork collapse have the potential to lead to loss of heterozy-

gosity (LOH). Indeed, in both human precancerous lesions and in a human skin xenograft hyperplasia model there is LOH that occurs predominantly at the common fragile sites, invoking the presence of DNA replication stress (3, 4). Expression of oncogenes in nontransformed cells also leads to DNA replication stress. Such cells have prematurely terminated DNA replication forks, DNA DSBs that form specifically in S phase, stretches of single-stranded DNA, and LOH targeting preferentially the common fragile sites (10, 11).

A key prediction of the oncogene-induced DNA damage model is that genomic instability in human precancerous lesions and in cancer is induced by the oncogenes themselves and is a feature of cancer from its earliest stages. How does this prediction fit with our current understanding of the mechanisms leading to genomic instability in human cancer? In advanced cancers, multiple mechanisms, including hypoxia, inflammation, and cell-matrix detachment, may contribute to genomic instability, but in precancerous lesions fewer mechanisms are likely to be implicated.

One mechanism involves telomere erosion, which can lead to transient surges in genomic instability, when telomeres become critically short but before telomerase expression is induced (5). Telomere erosion almost certainly contributes to genomic instability in human cancer, but whether it contributes to genomic instability in human



**Fig. 2.** Oncogene-induced DNA damage model for cancer development and progression. Genomic instability and tumor suppression are direct outcomes of oncogene-induced DNA replication stress and are both present from the beginning of cancer development, before the transition from precancerous lesion to cancer.

precancerous lesions is less clear. Telomere erosion is common in human precancerous lesions (6), but human skin xenografts already show genomic instability at common fragile sites a few weeks after induction of hyperplasia, before any changes in telomere length become evident (3).

A second mechanism involves permanent increases in genomic instability induced by mutations in genes whose function is to preserve genomic integrity. These so-called caretaker genes include DNA repair and cell cycle checkpoint genes (7). Indeed, one form of genomic instability, referred to as microsatellite instability (MIN), is caused by mutations in DNA mismatch repair genes (8). However, most human cancers do not have MIN, but rather have chromosomal instability (CIN), and the caretaker genes, whose deregulation is responsible for CIN, have not been identified (24). *p53* was considered initially a "guardian of the genome," but *p53*-null mice do not have overt genomic instability (25). In experimental models, candidate caretaker genes have been identified, but none of them are frequently mutated in sporadic human cancers (25). Further, large-scale sequencing analysis of all known coding regions in breast and colorectal cancers has failed to identify other genes that could account for CIN in a large fraction of human cancers (26). Because DNA DSBs can lead to CIN, it is possible that oncogene-induced DNA damage, rather than inactivation of caretaker genes, may contribute to CIN in human cancer, a hypothesis that could also explain why, in cell fusion experiments, CIN, unlike MIN, is a dominant trait (8).

Genomic instability is considered critical for cancer development because it would be difficult otherwise for a normal cell to accumulate all the mutations necessary to become a cancer cell (27). It is still unclear whether genomic instability induced by DNA replication stress has any role in cancer progression, but, in oral precancerous lesions, LOH at the common fragile site FRA3B predicted progression to cancer much

better than any other marker, including *p53* mutations and LOH at the *INK4a/arf* locus (28).

### Unanswered Questions and Future Prospects

An oncogene-induced DNA damage model for cancer development and progression can help explain many features of cancer, including its tendency to progress, the presence of genomic instability, and the tumor suppressor role of DNA DSB checkpoint proteins such as *p53* (Fig. 2). Of course, some aspects of the model need to be better defined. What is the mechanism by which oncogenes induce DNA replication stress? Under what conditions do oncogenes induce DNA replication stress, and is the strength of the oncogenic signal a critical factor, as suggested (29)? What is the role of ARF in human cancer development? What fraction of genomic instability is due to DNA replication stress and what fraction to other factors, such as telomere erosion? Does this model apply to most cancers, or are there tumor types, perhaps those characterized by low levels of genomic instability, in which this model does not apply? Lastly, does this model suggest new ways to diagnose and treat cancer?

With regard to the last question, one can envision new diagnostic methods for precancerous lesions. LOH at the common fragile site FRA3B shows promise in identifying high-risk precancerous lesions (28). It may also be possible to establish immunohistochemistry assays for DNA damage responses in a clinical setting on the basis of the assays currently available for research (3, 4, 13). An interesting possibility would be to develop cancer therapies, capitalizing on the presence of DNA replication stress specifically in cancer cells. DNA replication needs to resume after replication fork collapse. We suspect that cancer cells would not fare well if resumption of DNA replication could be inhibited.

### References and Notes

1. L. B. Schultz *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **65**, 489 (2000).

2. R. A. DiTullio Jr. *et al.*, *Nat. Cell Biol.* **4**, 998 (2002).
3. V. G. Gorgoulis *et al.*, *Nature* **434**, 907 (2005).
4. J. Bartkova *et al.*, *Nature* **434**, 864 (2005).
5. R. S. Maser, R. A. DePinho, *Science* **297**, 565 (2002).
6. D. E. Hansel *et al.*, *Mod. Pathol.* **19**, 772 (2006).
7. L. H. Hartwell, M. B. Kastan, *Science* **266**, 1821 (1994).
8. C. Lengauer, K. W. Kinzler, B. Vogelstein, *Nature* **396**, 643 (1998).
9. N. C. Denko, A. J. Giaccia, J. R. Stringer, P. J. Stambrook, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5124 (1994).
10. J. Bartkova *et al.*, *Nature* **444**, 633 (2006).
11. R. Di Micco *et al.*, *Nature* **444**, 638 (2006).
12. M. B. Kastan, J. Bartek, *Nature* **432**, 316 (2004).
13. J. Campisi, *Science* **309**, 886 (2005).
14. D. Hanahan, R. A. Weinberg, *Cell* **100**, 57 (2000).
15. C. J. Sherr *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **70**, 129 (2005).
16. C. Greenman *et al.*, *Nature* **446**, 153 (2007).
17. D. W. Bell *et al.*, *Science* **286**, 2528 (1999).
18. C. H. Westphal *et al.*, *Nat. Genet.* **16**, 397 (1997).
19. L. Moore *et al.*, *Oncogene* **22**, 7831 (2003).
20. M. A. Christophorou, I. Ringshausen, A. J. Finch, L. B. Swigart, G. I. Evan, *Nature* **443**, 214 (2006).
21. S. Ortega, M. Malumbres, M. Barbacid, *Biochim. Biophys. Acta* **1602**, 73 (2002).
22. A. Lengronne, E. Schwob, *Mol. Cell* **9**, 1067 (2002).
23. M. F. Arlt, S. G. Durkin, R. L. Ragland, T. W. Glover, *DNA Repair (Amsterdam)* **5**, 1126 (2006).
24. H. Rajagopalan, C. Lengauer, *Nature* **432**, 338 (2004).
25. D. P. Lane, *Nature* **358**, 15 (1992).
26. L. D. Wood *et al.*, *Science* **318**, 1108 (2007); published online 11 October 2007 (10.1126/science.1145720).
27. R. A. Beckman, L. A. Loeb, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 14140 (2006).
28. L. Mao *et al.*, *Nat. Med.* **2**, 682 (1996).
29. C. J. Sarkisian *et al.*, *Nat. Cell Biol.* **9**, 493 (2007).
30. We apologize for not being able to include all relevant concepts and references. We thank all our collaborators for their contributions to the development of the oncogene-induced DNA damage model. The authors are supported by grants from the Swiss National Foundation and the U.S. National Cancer Institute (T.D.H.); the Greek General Secretariat for Research and Technology Program to Enhance Research (V.G.G.); and the Danish Cancer Society, the Danish National Research Foundation, the Czech Ministry of Education (MSM6198959216), and the European Commission "Active p53" and "Mutant p53" Integrated Projects (J.B.).

### Supporting Online Material

[www.sciencemag.org/cgi/content/full/319/5868/1352/DC1](http://www.sciencemag.org/cgi/content/full/319/5868/1352/DC1)  
References

10.1126/science.1140735



# Dimethylsulfoniopropionate as a Foraging Cue for Reef Fishes

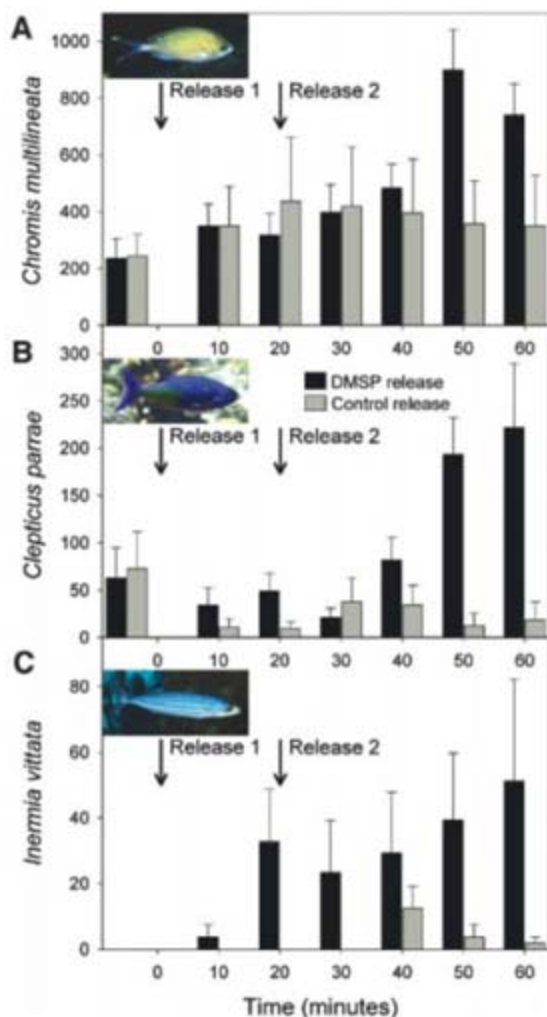
Jennifer L. DeBose,<sup>1</sup> Sean C. Lema,<sup>2</sup> Gabrielle A. Nevitt<sup>1\*</sup>

Coral reefs resemble islands of productive habitats where fishes aggregate, forage, and spawn. Although it has been suggested that some reef fishes use biogenic compounds as aggregation cues, specific compounds have not been identified. Dimethylsulfoniopropionate (DMSP) is produced by phytoplankton and by benthic algae associated with coral reefs and is linked to ocean productivity (1). DMSP is released during grazing by zooplankton or when herbivores are eaten (2, 3), suggesting a role as a foraging cue. DMSP has been studied intensively for its role in oceanic sulfur cycles and global climate regulation, but its ecological importance to marine fishes is unknown. Here, we present evidence that planktivorous reef fishes will aggregate to controlled experimental deployments of DMSP over coral reef habitats in the wild.

Release experiments were conducted at four sites on fringing reefs along 62 km of coastline off Curaçao, Netherlands Antilles. Sites were separated by at least 14 km. Our aim was to produce a down-current DMSP plume approximating concentrations ( $10^{-7}$  to  $10^{-9}$  M) that coral reef fish are likely to encounter in nature (4, 5). Two pairs of high-density polyethylene carboys (19 liters) containing either DMSP ( $10^{-7}$  M) or control (distilled water) were deployed 2 to 3 m apart at each reef site (6). Carboys were suspended midwater and anchored to the substrate. Carboys released DMSP or control solutions (0.5 liter per min) along the reef edge for 60 min, driven by the buoyant force of fresh water. We reasoned that, if fish used DMSP as a foraging cue, then more fish would be attracted to DMSP than to control plumes.

Results indicate that DMSP is a potent attractant to some planktivorous reef fish species, including brown chromis (*Chromis multilineata*, Fig. 1A, Friedman's test  $\chi^2 = 9.66$ ,  $P = 0.002$ ,  $n = 8$ ), Creole wrasse (*Clepticus parrae*, Fig. 1B,  $\chi^2 = 25.6$ ,  $P < 0.001$ ,  $n = 8$ ), and boga (*Inermia vittata*, Fig. 1C) (6). Fish also responded to DMSP following species-specific patterns. For example, brown chromis tend to associate with the reef crest but aggregate in the water column to forage on zooplankton (7). During DMSP releases, their numbers in the water column increased to nearly four times that of background levels. By contrast, Creole wrasse and boga tend to occur in open water or on the reef edge. These species recruited to the release

sites only in response to DMSP (Fig. 1, B and C). Additional analysis confirmed that there was no difference in background counts taken before release between release sites [analysis of variance (ANOVA), *Chromis*  $F_{3,8} = 3.43$ ,  $P = 0.072$ , and *Clepticus*  $F_{3,8} = 3.09$ ,  $P = 0.089$ ] or treatment



**Fig. 1.** Responses of coral reef fishes to DMSP and control releases over 1 hour. (A) *Chromis multilineata* (brown chromis), (B) *Clepticus parrae* (Creole wrasse), (C) *Inermia vittata* (boga). Fish counts plotted on the y axes are shown before release ("background" is to the left of 0 min) and at 10-min intervals during DMSP (black) and control (gray) releases. Arrows at 0 and 20 min indicate when the first and second sets of carboys were opened. Current velocity did not differ between DMSP and control releases (for DMSP, velocity =  $155 \pm 28$  cm/s; for control, velocity =  $213 \pm 66$  cm/s;  $t$  test,  $t = 0.64$ ,  $P = 0.55$ ), indicating that current velocity was not a strong determinant for the number of fish attracted. Note that y axes are shown at different scales. Data are expressed as mean  $\pm$  SEM ( $n = 8$ ). [Photos in (A) and (C) courtesy of F. and J. Burek]

groups (*Chromis*  $F_{1,8} = 0.009$ ,  $P = 0.925$  and *Clepticus*  $F_{1,8} = 0.047$ ,  $P = 0.833$ ).

The results are consistent with the foraging ecology of these species. Creole wrasse form midwater, off-reef foraging aggregations, whereas boga are open-water fishes that feed on zooplankton and only occasionally forage over reefs (7). Other planktivores such as sergeant major (*Abudefduf saxatilis*) and yellowtail snapper (*Ocyurus chrysurus*) did not respond to DMSP (for *Abudefduf*,  $\chi^2 = 0.118$ ,  $P = 0.731$ ,  $n = 8$ ; for *Ocyurus*,  $\chi^2 = 0.223$ ,  $P = 0.637$ ,  $n = 8$ ); however, they generally occurred in fewer numbers than the other species (average  $\pm$  SEM for *Abudefduf* was  $2.8 \pm 0.59$ ; for *Ocyurus*,  $1.0 \pm 0.18$ ) (6). These results imply that odors linked more closely to feeding activity than to the presence of prey alert some planktivorous fish species to potential foraging opportunities. These cues likely work in conjunction with visual cues provided by the foraging behavior of other fishes.

Although DMSP has been studied primarily in the context of global climate regulation, our results suggest that planktivorous fishes can eavesdrop on trophic interactions that leave residual chemical signatures and point to DMSP as a specific, biogenic compound that they can use to exploit their prey. Taken more broadly, this work strengthens the hypothesis that, in addition to serving as a climate regulator, DMSP is an important signal molecule in the marine environment (1). Recognizing this dual role should allow us to better understand and predict the effects of climate change on trophic interactions in the biological realm.

## References and Notes

1. M. Steinke, J. Stefels, E. Stams, *Limnol. Oceanogr.* **51**, 1925 (2006).
2. J. W. H. Dacey, G. M. King, P. S. Lobel, *Mar. Ecol. Prog. Ser.* **112**, 67 (1994).
3. H. Iida, *Bull. Tokai Reg. Fish. Res. Lab.* **124**, 35 (1988).
4. A. Broadbent, G. B. Jones, *Mar. Freshw. Res.* **55**, 849 (2004).
5. G. B. Jones, A. J. Trevena, *Mar. Freshw. Res.* **56**, 85 (2005).
6. Materials and methods are available on Science Online.
7. E. Lieske, R. Myers, *Coral Reef Fishes: Indo-Pacific and Caribbean Including the Red Sea* (Harper Collins, London, 1994).
8. Funded by the American Museum of Natural History and a PEO Scholar Award (to J.L.D.). We thank Caribbean Research and Management of Biodiversity Foundation (CARMABI), K. Bright, A. Gazit, B. Leysner, S. Strom, and R. Zimmer for logistical or technical assistance. Experiments adhered to Institutional Animal Care and Use Committee guidelines for animal research. Contribution number 2395 to Bodega Marine Laboratory.

## Supporting Online Material

[www.sciencemag.org/cgi/content/full/319/5868/1356/DC1](http://www.sciencemag.org/cgi/content/full/319/5868/1356/DC1)  
Materials and Methods  
References

28 September 2007; accepted 12 December 2007  
10.1126/science.1151109

<sup>1</sup>Neurobiology, Physiology, and Behavior and the Bodega Marine Laboratory, University of California, Davis, CA 95616, USA. <sup>2</sup>Biology and Marine Biology and the Center for Marine Science, University of North Carolina, Wilmington, NC 28403, USA.

\*To whom correspondence should be addressed. E-mail: [ganevitt@ucdavis.edu](mailto:ganevitt@ucdavis.edu)

# Long-Term Sea-Level Fluctuations Driven by Ocean Basin Dynamics

R. Dietmar Müller,<sup>1\*</sup> Maria Sdrolias,<sup>1</sup> Carmen Gaina,<sup>2</sup> Bernhard Steinberger,<sup>2</sup> Christian Heine<sup>1†</sup>

Earth's long-term sea-level history is characterized by widespread continental flooding in the Cretaceous period (~145 to 65 million years ago), followed by gradual regression of inland seas. However, published estimates of the Late Cretaceous sea-level high differ by half an order of magnitude, from ~40 to ~250 meters above the present level. The low estimate is based on the stratigraphy of the New Jersey margin. By assimilating marine geophysical data into reconstructions of ancient ocean basins, we model a Late Cretaceous sea level that is 170 (85 to 270) meters higher than it is today. We use a mantle convection model to suggest that New Jersey subsided by 105 to 180 meters in the past 70 million years because of North America's westward passage over the subducted Farallon plate. This mechanism reconciles New Jersey margin-based sea-level estimates with ocean basin reconstructions.

Long-term ( $10^7$  to  $10^8$  years) global (eustatic) sea-level fluctuations have been a driving force of biogeography, climate change, and organic evolution (1), yet they are poorly understood compared with short- to medium-term fluctuations ( $10^3$  to  $10^6$  years) (2). This is due to the diversity of potential driving factors involved, including changes in mid-ocean ridge length, spreading rates, oceanic area, sedimentation, mantle convection, superplumes, large igneous province emplacement, and ice volume (2). It is well documented that many continents, in particular North America, Europe, and Africa, were inundated by shallow seas peaking in the Late Cretaceous 80 million years ago (Ma), followed by a gradual sea-level fall (3), but the associated magnitude of global sea-level change is very controversial.

Subsidence of the sea floor away from mid-ocean ridges is caused by the thickening and cooling of a thermal boundary layer at Earth's surface (4). Most published investigations of long-term sea-level change rely on the analysis of the age-area and associated depth-area distribution of presently preserved ocean floor to derive an estimate of ocean depth and volume change through time, as pioneered by Pitman (5). Rowley (6) reviewed such reconstructions and concluded that substantial ocean basin volume change and associated sea-level change have not occurred since the breakup of Pangaea. In contrast, Kominz (7) partially reconstructed past mid-ocean ridges (isochrons) for the last 80 million years and estimated the sea level to have been 230 +135/-185 m at 80 Ma, and Xu *et al.*

(8), following a similar methodology, concluded that crustal production rates have decreased by 20 to 30% since 65 Ma, accompanied by a lowering of sea level between 125 and 250 m. Haq *et al.* (9) estimated the magnitude of long-term sea-level change from the present elevation of ancient marine deposits after reconstructing their subsidence history and adopting a sea-level tiepoint of 242 m above the present level for 80 Ma from Harrison (10), primarily on the basis of Kominz's (7) mid-ocean ridge volume analysis. In contrast, long-term sea-level estimates from stratigraphic sections of the New Jersey margin of North America yield a substantially lower 40-m sea-level high at 80 Ma (2), building on the work of Watts and co-workers (11, 12), who suggested that global sea-level signals can be isolated from the North American east coast sedimentary record.

We present a comprehensive reconstruction of the global age-area and depth-area distribution of ocean floor, including remnants of subducted crust, since the Early Cretaceous (140 Ma), to compute the effects of changes in crustal production, sediment thickness, and ocean-basin depth and area on sea-level fluctuations through time. In addition, we use a mantle convection model to test Miller *et al.*'s (2) sea-level curve and its premise that the New Jersey shelf has not been affected by tectonic processes other than thermal subsidence since the Cretaceous. This assumption allowed Miller *et al.* (2) to argue that the regional sea-level curve from New Jersey margin stratigraphy represents global (eustatic) sea-level variations.

**Reconstructing vanished oceans.** We establish the locations and geometry of mid-ocean ridges through time on the basis of marine magnetic anomaly identifications; geological information such as paleomagnetic data from terranes and microcontinents, especially in the Tethys Ocean (13); mid-oceanic ridge subduction events; and the rules of plate tectonics (14). On

the basis of a global set of tectonic plate rotations (15, 16), we construct a set of refined sea-floor isochrons (Fig. 1) following the interpolation technique outlined by Müller *et al.* (15, 16) but including a multitude of additional data (see Supporting Online Material).

In areas where one flank of a mid-ocean ridge system has been subducted, we assume spreading symmetry to reconstruct the subducted flank. This is a reasonable assumption given that globally, the maximum cumulative spreading asymmetry has been found to be less than 10% (17). The now entirely subducted ocean floor north of Africa, India, and Australia (the Meso-Tethys, Fig. 1) was reconstructed with the model of Heine *et al.* (18). Sea-floor isochrons for the Proto-Caribbean were constructed on the basis of the Early Cretaceous divergence history of the Americas, followed by the eastward insertion of the Farallon Plate into the Caribbean, leading to subduction of Proto-Caribbean ocean crust in the Late Cretaceous (19). Our Cretaceous reconstruction of the Panthalassa Ocean (Fig. 1) is based on the premise that the Pacific Plate formed at 170 Ma in the Late Jurassic as a triangle, originating from a triple junction between the Farallon, Phoenix, and Izanagi plates (20), which are reconstructed on the basis of preserved magnetic lineations and fracture zones (21, 22) and the assumption of spreading symmetry, with the exception of areas where ridge jumps are documented. We reconstruct the now entirely subducted Izanagi Plate and its conjugate Pacific Ocean floor (Fig. 1) following Whittaker *et al.* (23). Locations of plate boundaries within the Pacific/Panthalassa Ocean relative to surrounding continents are reconstructed with a plate circuit through Antarctica after 83.5 Ma and with Pacific and African plate motion relative to hot-spots independently determined for earlier times (15, 16).

Plate boundary reconstructions between the Meso-Tethys and Panthalassa oceans are based on modeled motion vectors between the Izanagi, Phoenix, Australian, and Meso-Tethys plates. We use the combined model of finite rotations, sea-floor isochrons, mid-ocean ridge locations, and outlines of boundaries between continental and oceanic crust to construct a set of oceanic paleo-age grids from 140 Ma to the present (Fig. 1 and fig. S1), with accompanying age uncertainty grids (fig. S2). The mean error of our modeled age-area distribution of today's ocean floor is about 2.8 million years. We scale our error analysis for ocean basin reconstructions such that the mean reconstruction error of now subducted ocean floor is about 10 million years, namely, a fourfold increase in the combined uncertainties reflecting errors in mid-ocean ridge location and geometry.

**Oceanic depth through time.** An analysis of crustal production through time in 5-million-year intervals (fig. S3) illustrates that the most severe

<sup>1</sup>EarthByte Group, School of Geosciences, Building H11, The University of Sydney, Sydney, NSW 2006, Australia.

<sup>2</sup>Center for Geodynamics, Geological Survey of Norway, Leiv Eirikssons vei 39, N-7491 Trondheim, Norway.

\*To whom correspondence should be addressed. E-mail: dietmar@geosci.usyd.edu.au

†Present address: StatoilHydro, Global Exploration Technology, Drammensveien 264, N-0240 Oslo, Norway.

fluctuations in spreading-ridge crustal production are recorded in the Pacific Ocean. The largest drop in global crustal production, of about 50%, is recorded from 65 to 60 Ma, corresponding to the subduction of the Izanagi-Pacific spreading ridge (23) (Fig. 1 and fig. S1). This event, together with subsequent ridge subduction episodes along the Americas (fig. S1), leads to a mean Pacific sea-floor age difference of 35 million years (My) between 120 Ma and today (excluding back-arc basins), whereas all oceanic crust outside the Pacific Ocean taken together has experienced a total mean age difference of about 10 My (Fig. 2A). This demonstrates that dramatic changes in the age-area distribution of ocean floor have been driven primarily by cyclic changes in mid-ocean ridge creation, evolution, and destruction in the Pacific Ocean. It also shows that the use of the present distribution of the ocean-floor ages to obtain paleo-age distributions without reconstructing subducted sea floor (6, 24, 25) is not valid.

The maximum global difference in mean oceanic crustal age between 120 Ma and the present is about 26 My, ranging from  $38.4 \pm 4.1$  My at 120 Ma to  $64.1 \pm 1.4$  My at present (Fig. 2B). This is in contrast to the 6-My difference predicted by Cogne *et al.* (25), who erroneously assumed that the Pacific age-area distribution has remained constant since the breakup of Pangaea. Using a published age-depth relation (GDH-1) (26), we compute the depth-area distribution of the ocean basins (Fig. 2C). We choose GDH-1 for converting age to depth because this relation is based on sediment-corrected depths without excluding data from hotspot swells and seamounts. Therefore, GDH-1 provides a good average fit to sediment-unloaded oceanic basement

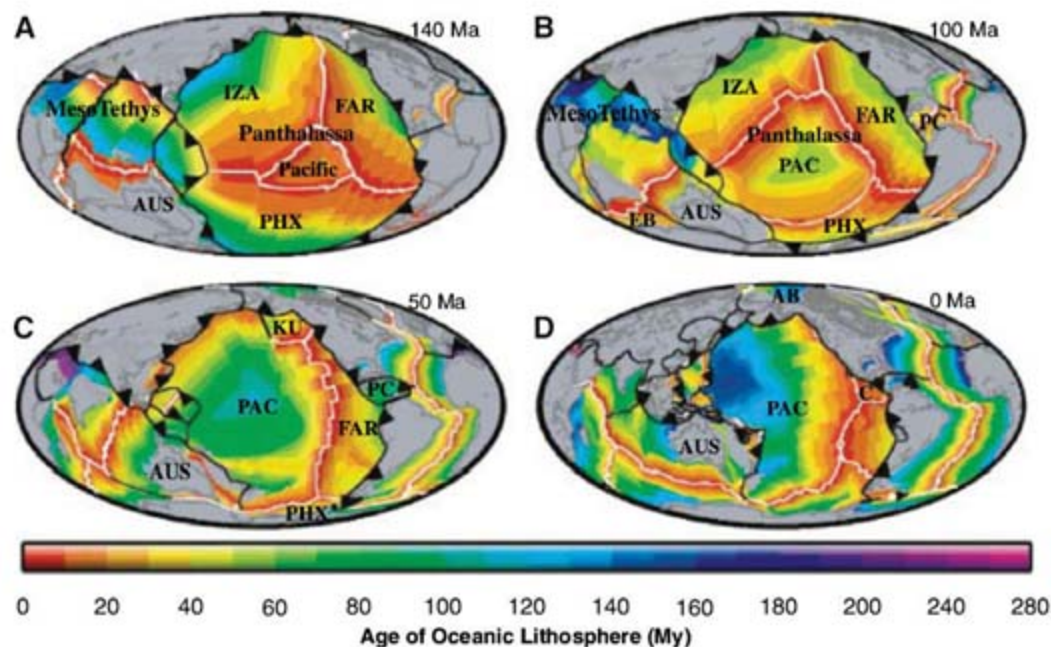
depths (26) and is preferable for predicting the average oceanic basement depth through time, including thermally rejuvenated lithosphere, as compared to models [e.g., (27, 28)] that reflect ocean-depth changes related to plate aging through time only.

Three additional factors play an important role in controlling global ocean basin depth through time, namely, the generation of oceanic large igneous provinces (LIPs), oceanic sedimentation (2), and changes in oceanic crustal area. We use Schubert and Sandwell's (29) method to determine the average elevation of individual major oceanic plateaus (table S1) relative to the surrounding sea floor, based on the difference between the modal depths within two polygons, one outlining the perimeter of a given plateau from a LIP database (30) and another polygon including surrounding ocean floor. Plateaus are added to our basement depth grids at their emplacement time based on a revised compilation of eruption ages (31). We estimate the uncertainty in LIPs-related mean oceanic depth changes to be on the order of up to  $\pm 5$  m. An additional effect not considered here is the potential uplift of the sea floor some time before LIP eruption when the body of hot material (plume head) arrives in the upper mantle, followed by a collapse during eruption. From estimates for the size (about 500 km in diameter) and density anomaly (about 1%) of the plume head, we infer about a 2-m sea-level rise before eruption of a LIP, with 5 m probably an upper limit (corresponding to  $\sim 700$  km in diameter).

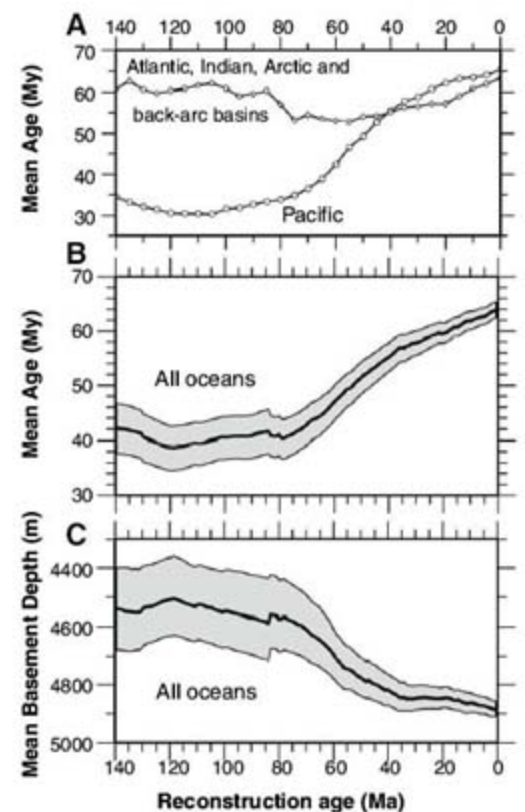
As the ocean floor ages, its sediment cover thickens, but abyssal sediment thickness is also latitude dependent, as illustrated in a polynomial surface fit of global sediment thickness (32) as a

function of oceanic crustal age and latitude (fig. S4). We use this model to estimate the distribution of total oceanic sediment thickness and its isostatic compensation (33) through time, associated with an estimated mean sediment thickness uncertainty of about  $\pm 15$  m for the Tertiary and  $\pm 25$  m for the Cretaceous; however, this uncertainty is not well constrained. We create a set of paleo-bathymetry maps by adding major oceanic plateaus (table S1) and sediment thickness to our reconstructed basement depth maps (Fig. 3). These maps allow us to compute oceanic crustal area and mean depth through time.

The emplacement of LIPs alone has resulted in a total sea-level rise of nearly 100 m since 140 Ma, with major pulses at about 120, 110, 90, 60, and 40 Ma (Fig. 4 and fig. S6). The sediment thickness effect on sea level through time is estimated to be an 18-m drop in sea level from 140 to 82 Ma, followed by a sea-level rise of 61 m (fig. S6). Consequently, the modeled increase in



**Fig. 1.** Age-area distribution of the ocean floor at (A) 140 Ma, (B) 100 Ma, (C) 50 Ma, and (D) the present day. See fig. S1 for paleo-age-area distribution in 10-My intervals. PAC, Pacific Plate; FAR, Farallon Plate; PHX, Phoenix Plate; IZA, Izanagi Plate; KU, Kula Plate; AUS, Australian Plate; PC, Proto-Caribbean; EB, Enderby Basin; AB, Amerasian Basin; C, Cocos Plate. Active mid-ocean ridges are represented as white lines, and subduction locations as black lines and triangle symbols.



**Fig. 2.** (A) Mean oceanic crustal age of the Pacific versus Indian, Atlantic, and Arctic ocean basins including all back-arc basins. The Pacific Ocean mean age increased from 30 to 65 My between 120 Ma and the present, whereas the mean age of all other ocean basins changed relatively little in comparison (a maximum of 10 My), challenging the notion that the Pacific Ocean mean age has remained constant through time (25). (B) Mean oceanic crustal age in 1-My intervals from 140 Ma to the present with error envelope computed from the age error grids (see fig. S2 for age error grids in 10-My intervals). (C) Mean oceanic basement depth since 140 Ma with uncertainties derived from age errors based on the GDH-1 age-depth model (26).

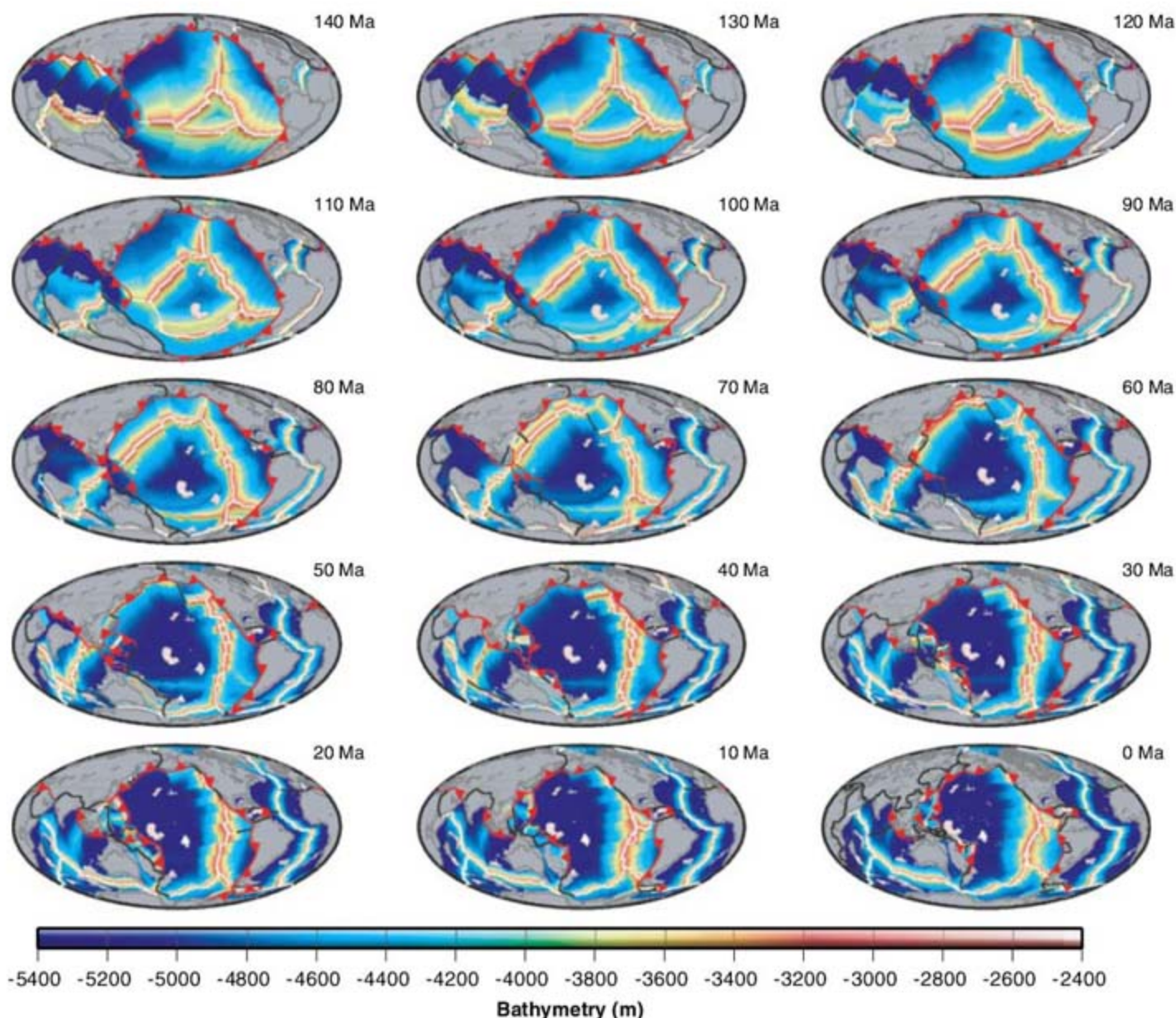
mean basement depth of 350 m from 120 Ma to the present (Fig. 2C) is alleviated by oceanic plateau emplacement and sediment thickness fluctuations to a net depth increase of about 260 (+100, -85) m since 118 Ma (fig. S6).

**Long-term sea-level change.** Global sea-level change during the past 33 My was dominated by large ice-sheet growth and decay (glacio-eustasy) (2). As our sea-level curve does not include the effects of glacial ice, we shift it by the 54 m that would be added to sea level if all present ice sheets melted (2) (Fig. 4), resulting in a Cretaceous high of 304 m at 119 Ma. Changes in sea-level  $\Delta s_l$  as they would appear to an observer on land are related to changes in mean ocean basement depth  $\Delta h$  via isostatic compensation (34, 35):  $\Delta s_l = (\rho_m - \rho_w) \rho_m \times \Delta h$  ( $\rho_m$  is mantle density,  $3300 \text{ kg m}^{-3}$ ;  $\rho_w$  is water density), yielding an apparent sea-level drop of about 235 m since a maximum at 118 Ma (fig. S6). However, this estimate, based purely on mean oceanic depth change, does not account for

changes in oceanic area. Our reconstructions suggest that since 140 Ma, the global oceanic area has decreased by 3.3%, corresponding to a loss of about 1 million  $\text{km}^2$ . This is due to continental crustal stretching during rifting, preceding continental breakup and ocean basin formation. Assuming no change in global oceanic water volume, the modeled decrease in oceanic area through time partly counteracts the aging and associated deepening of the global oceans after 80 Ma and attenuates the modeled Late Cretaceous sea-level high to about 170 m (fig. S6). Continental margin thermal subsidence following continental stretching is compensated mostly by sedimentation that keeps continental shelves relatively close to sea level (36), resulting in small changes in ocean basin depth from passive margin aging, which is not considered here.

A sea-level curve (Fig. 4) that takes into account both changes in depth and area of the global ocean basins reduces the modeled sea-level high at 118 Ma to about 150 m (Fig. 4). We

obtain a total sea-level rise of about 70 m between 140 and 110 Ma; this is a better match with most Cretaceous continental flooding estimates [summarized by (2)] than our alternative sea-level curve not considering changes in oceanic area, which implies a sea-level rise of less than 30 m for the same time period (fig. S6). We model a long-term sea-level fall measuring about 100 m from 80 to 30 Ma, driven by the aging and deepening of the global ocean floor during this time (Fig. 2 and fig. S6). Our Late Cretaceous sea-level maximum of 170 m (90 to 265 m) occurs at 82 Ma, close to the benchmark time of 80 Ma (2, 7), supporting estimates from continental inundation as summarized by Miller (2) and close to Bond's (37) estimate of 200 m as well as that by Watts and co-workers (11, 12), but differing considerably from Miller *et al.*'s (2) best sea-level estimate of  $\sim 40$  m at 80 Ma (Fig. 4), based on New Jersey margin stratigraphy, raising the question of what causes this substantial disagreement.



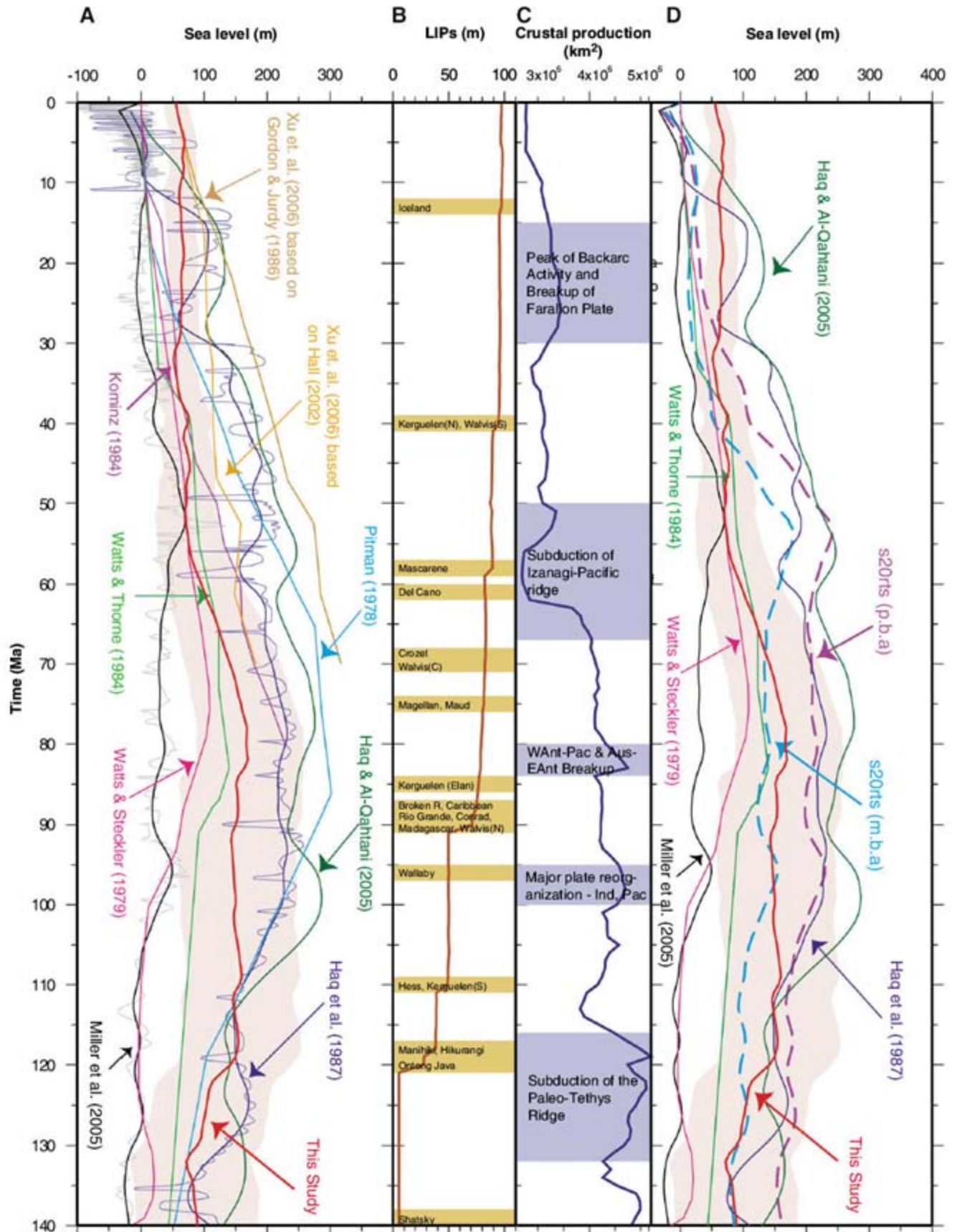
**Fig. 3.** Reconstructed bathymetry through time by combining basement depths derived from paleo-age grids (Fig. 1 and fig. S1) and the GDH-1 age-depth model (26) with estimates of sediment thickness through time (figs. S4 and S5) and inclusion of all major oceanic plateaus (table S1).

**New Jersey margin subsidence.** We propose that the large 130-m discrepancy at 80 Ma between our sea-level curve and that from Miller *et al.* (2) (Fig. 4) reflects the progressive, mantle-driven tectonic subsidence of the New Jersey margin over the negatively buoyant Farallon

slab imaged in the lower mantle (Fig. 5), while global sea level was falling after the Late Cretaceous highstand. Our reconstructions of subduction trench positions through time (Fig. 5A) suggest that the east coast of North America has moved over the Farallon slab (Fig. 5B)

during the past 60 My. We test this hypothesis using regional results from a well-established global mantle convection modeling approach (38) to compute time-dependent dynamic surface topography by advecting present mantle density anomalies, derived from seismic mantle

**Fig. 4.** (A) Predicted relative sea-level change (red) and error envelope (reflecting basement age-depth errors) (light pink) since 140 Ma based on changes in the average depth and area of the ocean basins using the GDH-1 plate model (26) age-depth conversion (Database S1) compared with the sea-level curves by Kominz (7) (magenta), Haq *et al.* (9) (dark blue), adjusted to the Berggren *et al.* time scale (2), Haq and Al-Qahtani (43) (dark green), Miller *et al.* (2) (black), Watts and Steckler (11) (pink), Watts and Thome (12) (light green), Pitman (5) (light blue), and the two curves by Xu *et al.* (8) (light and dark orange). The Haq *et al.* (9), Haq and Al-Qahtani (43), and Miller *et al.* (2) curves have been smoothed by a 10-My cosine-arch low-pass filter (in bold), representing long-term sea-level change. The original curves from Haq *et al.* (9) and from Miller *et al.* (2) are shown as thin lines. (B) Sea-level effect of large igneous provinces (LIPs) (100 m cumulatively). (C) Total crustal production through time since 140 Ma with major tectonic events highlighted in blue. (D) Miller *et al.* (2) sea-level curve, uncorrected (black) and corrected for post-70 Ma modeled tectonic subsidence of the New Jersey margin based on the *s20rts* seismic tomography model (39) for the pure backward advection (p.b.a) (purple) and modified backward advection (m.b.a) (light blue) compared with long-term sea-level variation from our computation of the volume of the ocean basins through time (red), as well as with the models by Haq and co-workers (9, 43) and Watts and co-workers (11, 12).



Miller *et al.* (2) sea-level curve, uncorrected (black) and corrected for post-70 Ma modeled tectonic subsidence of the New Jersey margin based on the *s20rts* seismic tomography model (39) for the pure backward advection (p.b.a) (purple) and modified backward advection (m.b.a) (light blue) compared with long-term sea-level variation from our computation of the volume of the ocean basins through time (red), as well as with the models by Haq and co-workers (9, 43) and Watts and co-workers (11, 12).

tomography, backward in the mantle flow field, which in turn is computed from mantle density anomalies and given surface plate velocities. We model the flow field of the mantle through time using a spectral method based on spherical harmonic expansion of surface plate velocities and internal density heterogeneities at each depth level (38). Dynamic topography is computed beneath air with a free upper boundary and the same viscosity structure as in (38) with a high-viscosity lithosphere ( $2.4 \times 10^{22}$  Pa·s). We compare regional time-dependent dynamic topography since 70 Ma for the New Jersey coast of North America on the basis of mantle density anomalies derived for the present day from three commonly used seismic tomography models, *s20rts*, *smean*, and *ngrand* (39–41), using a conversion factor of 0.25 from relative seismic velocity to relative density variations below 220 km and disregarding velocity variations above 220 km. In addition to the pure backward advection previously used (38), we also employ a modified procedure to reconstruct past mantle density anomalies. It differs in two ways: (i) Backward-advected upwellings are always continued up to 220 km at each time step, and thus the restoration of past mantle upwellings is improved; (ii) backward-advected downwellings are removed in the uppermost 220 km, and thus the effect of reconstructed

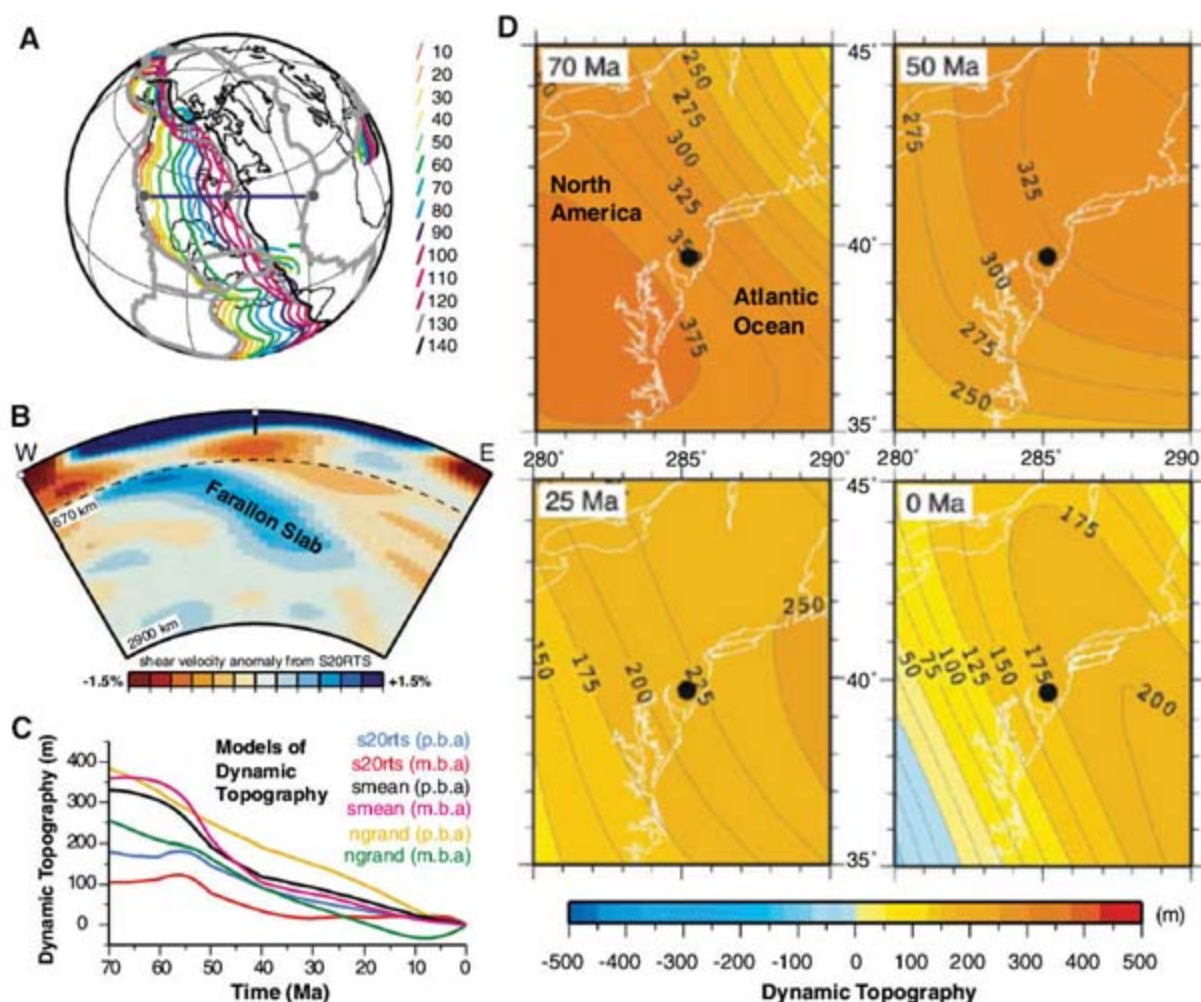
ocean floor, which is already explicitly included, is not counted twice.

All models result in New Jersey tectonic subsidence during the past 70 million years, with amplitudes of 105 to 180 m (*s20rts*), 330 to 360 m (*smean*), and 255 to 385 m (*ngrand*), illustrating differences in the seismic tomography models and resulting model uncertainties (Fig. 5C). Mantle flow models tend to overpredict dynamic topography compared to estimates based on observed topography (42). However, all models agree with the qualitative consideration of the New Jersey margin overriding the Farallon slab, with *s20rts* providing the most reasonable estimates. Without taking dynamic topography into consideration, Miller *et al.*'s (2) model leads to an underestimate of the post-Late Cretaceous global sea-level drop. Therefore, we add the *s20rts*-based modeled New Jersey margin mantle-driven subsidence to Miller *et al.*'s (2) sea-level curve (Fig. 4D) and assume that before 70 Ma, prior to overriding the Farallon slab, the New Jersey margin was not affected by changes in dynamic topography. Miller *et al.*'s (2) corrected sea-level curves based on the *s20rts* seismic tomography models are both largely within the error envelope of our Cretaceous sea-level curve based on ocean basin volume analysis (Fig. 4D), implying a Late Cretaceous sea-level high between 142 and 217 m (based on *s20rts*) at 60

Ma. In particular, the modified backward advection *s20rts*-based tectonic subsidence correction brings Miller *et al.*'s (2) curve in good alignment with the Cretaceous portion of our ocean basin volume-derived curve (Fig. 4D), whereas the pure backward advection (38) *s20rts*-based corrected Miller *et al.*'s (2) curve is similar in amplitude to Haq *et al.*'s curve (9, 43).

Our results support the idea that mantle convection-driven dynamic topography has played an important role in the subsidence history of the New Jersey margin, predicting a total post-70 Ma tectonic subsidence between 105 and 180 m at average rates of 1.5 to 2.5 m/My. For times after 110 Ma, our ocean basin volume-derived sea-level curve yields lower estimates than those by Haq and co-workers (9, 43). The discrepancies may reflect global effects of dynamic topography, not modeled here, because the mantle backward advection that we use cannot realistically restore all global mantle density anomalies in the past. However, a combination of a global ocean basin volume analysis with modern geodynamic models provides a powerful tool for discriminating eustasy from regional, time-dependent sea-level variations caused by mantle convection. Our global sea-level curve calibration provides an improved framework for sequence stratigraphy, resource exploration, and models for long-term climate change.

**Fig. 5.** (A) Reconstructed subduction zone locations in present-day coordinates with coastlines in black and present-day plate boundaries in gray with ages of trench positions color coded. Legend numbers indicate reconstructed ages (in million years) in an absolute mantle reference frame (see text). Horizontal bold blue line and gray circles outline location of mantle seismic tomography profile below. (B) West-east mantle seismic tomography cross-section (39) centered on the New Jersey margin [see (A) for profile location], imaging the subducted Farallon slab in the lower mantle underneath the margin. Our kinematic model predicts North America's east coast to be underlain by subducted slabs about 80 to 140 million years old, coinciding with the location of the imaged Farallon slab (39–41). (C) Predicted dynamic topography of the New Jersey margin from 70 Ma to the present based on the *s20rts*, *smean*, and *ngrand* seismic tomography models for the pure (p.b.a) and modified (m.b.a) backward advection cases (39–41) (table S2). (D) Modeled dynamic topography of eastern North America at 70, 50, 25 Ma and at present based on the *s20rts* (p.b.a) seismic tomography model, in the reference frame of the North American plate. The black dot outlines the location of the New Jersey margin.



## References and Notes

- J. D. Hays, W. C. I. Pitman, *Nature* **246**, 12 (1973).
- K. G. Miller *et al.*, *Science* **310**, 1293 (2005).
- A. Hallam, *Am. J. Sci.* **261**, 397 (1963).
- D. L. Turcotte, E. R. Oxburgh, *J. Fluid Mech.* **28**, 29 (1967).
- W. C. Pitman III, *Geol. Soc. Am. Bull.* **89**, 1389 (1978).
- D. B. Rowley, *Geol. Soc. Am. Bull.* **114**, 927 (2002).
- M. A. Kominz, in *Interregional Unconformities and Hydrocarbon Accumulation*, J. S. Schlee, Ed. (American Association of Petroleum Geologists, Tulsa, OK, 1984), vol. 36, pp. 109–127.
- X. Q. Xu, C. Lithgow-Bertelloni, C. P. Conrad, *Earth Planet. Sci. Lett.* **243**, 552 (2006).
- B. U. Haq, J. Hardenbol, P. R. Vail, *Science* **235**, 1156 (1987).
- C. G. A. Harrison, in *Sea-Level Change*, R. R. Revelle, Ed. (National Academy Press, Washington, DC, 1990), pp. 141–158.
- A. B. Watts, M. S. Steckler, in *Deep Sea Drilling Results in the Atlantic Ocean: Continental Margin and Paleoenvironment*, M. Talwani, Ed. (American Geophysical Union, Washington, DC, 1979), vol. 3, pp. 235–248.
- A. B. Watts, J. A. Thorne, *Mar. Petrol. Geol.* **1**, 319 (1984).
- G. M. Stampfli, G. D. Borel, *Earth Planet. Sci. Lett.* **196**, 17 (2002).
- A. Cox, B. R. Hart, *Plate Tectonics: How It Works* (Blackwell Science, Oxford, 1986).
- R. D. Müller, W. R. Roest, J.-Y. Royer, L. M. Gahagan, J. G. Sclater, *J. Geophys. Res.* **102**, 3211 (1997).
- R. D. Müller, M. Sdrolias, C. Gaina, W. R. Roest, *Geochem. Geophys. Geosyst.* 10.1029/2007GC001743 (2008); available at [www.agu.org/journals/gc/](http://www.agu.org/journals/gc/).
- R. D. Müller, W. R. Roest, J.-Y. Royer, *Nature* **396**, 455 (1998).
- C. Heine, R. D. Müller, C. Gaina, in *Continent-Ocean Interactions Within East Asian Marginal Seas*, AGU Monograph 149, P. D. Clift, D. E. Hayes, W. Kuhnt, P. Wang, Eds. (American Geophysical Union, Washington, DC, 2004), pp. 37–54.
- M. I. Ross, C. R. Scotese, *Tectonophysics* **155**, 139 (1988).
- D. C. Engebretson, A. Cox, R. G. Gordon, *Geol. Soc. Am. Spec. Pap.* **206**, 1 (1985).
- M. Nakanishi, K. Tamaki, K. Kobayashi, *Geophys. Res. Lett.* **19**, 693 (1992).
- T. Atwater, J. Severinghaus, in *The Eastern Pacific Ocean and Hawaii*, E. L. Winterer, D. M. Hussong, R. W. Decker, Eds. (Geological Society of America, Boulder, CO, 1990), vol. N, pp. 15–20.
- J. M. Whittaker *et al.*, *Science* **318**, 83 (2007).
- C. P. Conrad, C. Lithgow-Bertelloni, *Geology* **35**, 29 (2007).
- J. P. Cogne, E. Humler, V. Courtillot, *Earth Planet. Sci. Lett.* **245**, 115 (2006).
- C. A. Stein, S. Stein, *Nature* **359**, 123 (1992).
- A. G. Crosby, D. Mckenzie, J. G. Sclater, *Geophys. J. Int.* **166**, 553 (2006).
- M. P. Doin, L. Fleitout, *Geophys. J. Int.* **143**, 582 (2000).
- G. Schubert, D. Sandwell, *Earth Planet. Sci. Lett.* **92**, 234 (1989).
- M. F. Coffin, O. Eldholm, in *Magmatism and the Causes of Continental Breakup*, B. C. Storey, T. Alabaster, R. J. Pankhurst, Eds. (Geological Society of London, 1992), pp. 21–34.
- T. H. Torsvik, M. A. Smethurst, K. Burke, B. Steinberger, *Geophys. J. Int.* **167**, 1447 (2006).
- D. L. Divins (National Geophysical Data Center, 2004), [www.ngdc.noaa.gov/mgg/sedthick/sedthick.html](http://www.ngdc.noaa.gov/mgg/sedthick/sedthick.html).
- T. J. S. Sykes, *Mar. Geol.* **133**, 35 (1996).
- B. Parsons, *J. Geophys. Res.* **87**, 289 (1982).
- A sea-level rise  $\Delta s_l$  pushes the entire oceanic lithosphere down; thus, the resulting sea-level rise  $\Delta s_l$  is less than the changing mean ocean basement depth  $\Delta h$ , which is computed without this effect.
- P. L. Heller, C. L. Angevine, *Earth Planet. Sci. Lett.* **75**, 417 (1985).
- G. C. Bond, *Tectonophysics* **61**, 285 (1979).
- B. Steinberger, R. Sutherland, R. J. O'Connell, *Nature* **430**, 167 (2004).
- J. Ritsema, H.-J. van Heijst, J. H. Woodhouse, *J. Geophys. Res.* **109**, 10.1029/2003JB002610 (2004).
- T. W. Becker, L. Boschi, *Geochem. Geophys. Geosyst.* **3**, 48 (2002).
- S. P. Grand, *Philos. Trans. R. Soc. London Ser. A* **360**, 2475 (2002).
- B. Steinberger, *Phys. Earth Planet. Int.* **164**, 2 (2007).
- B. U. Haq, M. Al-Qahtani, *GeoArabia* **10**, 127 (2005).
- This paper benefited from reviews of early drafts by A. Dutkiewicz, J. Whittaker, and M. Gurnis, as well as from careful reviews by M. Kominz, R. O'Connell, and an anonymous reviewer, which improved the paper substantially. Ideas for New Jersey margin mantle-driven subsidence evolved during a sabbatical at the California Institute of Technology in discussions with M. Gurnis and S. Spasojevic. This work was funded by the Australian Research Council.

## Supporting Online Material

[www.sciencemag.org/cgi/content/full/319/5868/1357/DC1](http://www.sciencemag.org/cgi/content/full/319/5868/1357/DC1)  
Materials and Methods

Figs. S1 to S6

Tables S1 and S2

Database S1

9 October 2007; accepted 30 January 2008

10.1126/science.1151540

## Antisocial Punishment Across Societies

Benedikt Herrmann,<sup>1</sup> Christian Thöni,<sup>2</sup> Simon Gächter<sup>1\*</sup>

We document the widespread existence of antisocial punishment, that is, the sanctioning of people who behave prosocially. Our evidence comes from public goods experiments that we conducted in 16 comparable participant pools around the world. However, there is a huge cross-societal variation. Some participant pools punished the high contributors as much as they punished the low contributors, whereas in others people only punished low contributors. In some participant pools, antisocial punishment was strong enough to remove the cooperation-enhancing effect of punishment. We also show that weak norms of civic cooperation and the weakness of the rule of law in a country are significant predictors of antisocial punishment. Our results show that punishment opportunities are socially beneficial only if complemented by strong social norms of cooperation.

Recent research has shown that altruistic punishment, that is, a person's propensity to incur a cost in order to punish free-loaders who fail to pull their weight in cooperative endeavors, can explain why genetically unrelated individuals are often able to maintain high levels of socially beneficial cooperation (1–4). This holds even when direct and indirect reciprocity (5, 6) or laws and regulations provide no incentives to behave cooperatively (7).

In this paper, we direct attention to a phenomenon that [with a few exceptions (8–10)] has been largely neglected: People might punish not only free-loaders, but cooperators too. For example, participants who had been punished in the past for contributing too little might retaliate against the cooperators because the cooperators are precisely those individuals most likely to punish the free-riding low contributors. Our experimental evidence from 16 participant pools with various cultural and economic backgrounds shows that antisocial punishment of prosocial cooperators is indeed widespread in many participant pools; interestingly, the participant pools in which most of the previous research on altruistic punishment has been conducted form the main exception.

Our observation of antisocial punishment grew out of our research goal to understand whether there are cross-societal differences in people's punishment and cooperation behavior. Previous large-scale cross-cultural evidence comes mainly from one-shot bargaining games conducted in small-scale societies around the world (11, 12). However, there is no systematic large-scale evidence on cooperation games. We therefore conducted cooperation experiments with and without punishment opportunities. Moreover, we ran our experiments as repeated games to see whether different cooperation levels emerge and remain stable across groups. Such a possibility is precluded in one-shot experiments.

Our research strategy was to conduct the experiments with comparable social groups from complex developed societies with the widest possible range of cultural and economic backgrounds (13) to maximize chances of observing cross-societal differences in punishment and cooperation. The societies represented in our participant pools diverge strongly according to several widely used criteria developed by social scientists in order to characterize societies (14–16). This variation, covering a large range of the worldwide available values of the respective criteria, provides us with a novel test for seeing whether societal differences between complex societies have any impact on experimentally observable disparities in cooperation and punishment behavior.

**Experiments.** The workhorse for our cross-societal analysis is the public goods game with and without punishment (1). The public goods game is a stylized model of situations that require

<sup>1</sup>Centre for Decision Research and Experimental Economics, University of Nottingham, School of Economics, Sir Clive Granger Building, University Park, Nottingham NG7 2RD, UK. <sup>2</sup>University of St. Gallen, FEW-HSG, Vambuelstrasse 14, CH-9000 St. Gallen, Switzerland.

\*To whom correspondence should be addressed. E-mail: [simon.gaechter@nottingham.ac.uk](mailto:simon.gaechter@nottingham.ac.uk)

cooperation to achieve socially beneficial outcomes in the presence of free-rider incentives. Examples abound: warfare, cooperative hunting, voting, paying taxes, fighting corruption, contributing to public goods, teamwork, work morale, neighborhood watch, common pool resource management, recycling, tackling climate change, and so on. These are frequent situations with the common feature that cooperation leads to a group-beneficial outcome but is jeopardized by selfish incentives to ride free on others' contributions.

To implement a cooperation game with and without punishment opportunities, we adapted a design developed by (1). In each participant pool, we conducted the exact same public goods experiment with real monetary stakes and two treatment conditions: a no-punishment condition (the N experiment) and a punishment condition (the P experiment). Groups of four members played the following public goods game in both conditions: Each member received an endowment of 20 tokens. Participants had to decide how many tokens to keep for themselves and how many to contribute to a group project. Each of the four group members earned 0.4 tokens for each token invested in the project, regardless of whether he or she contributed any. Because the cost of contributing one token in the project was exactly one token whereas the return on that token was only 0.4 tokens, keeping all one's own tokens was always in any participant's material self-interest, irrespective of how much the other three group members contributed. Yet, if each group member retained all of his or her tokens, there were no earnings to be

shared; on the other hand, each member would earn  $0.4 \times 80 = 32$  tokens if each of them invested their entire 20-token endowment.

All the interactions in the experiment were computer-mediated (17) and took place anonymously. Participants were not informed about the identity of others in the group; they made their contribution decisions simultaneously, and, once the decisions were made, they were informed about the other group members' contributions.

The only and crucial difference between the P experiment and the N experiment was that participants in the P experiment could punish each of the other group members after they were informed about the others' investments, whereas the N experiment ended after participants were informed about the other group members' contributions. A punishment decision was implemented by assigning the punished member between zero and 10 deduction points. Each deduction point assigned reduced the punished member's earnings by three tokens and cost the punishing member one token. All punishment decisions were made simultaneously. Participants were not informed about who punished them.

One of the goals of our experiment was to see whether and at what level punishment stabilized cooperation in the P experiment compared to the N experiment. To allow for the emergence of different cooperation levels, we therefore repeated the experiment 10 times under both conditions, keeping the group composition constant.

Because we were interested in whether people behave differently under the exact same cir-

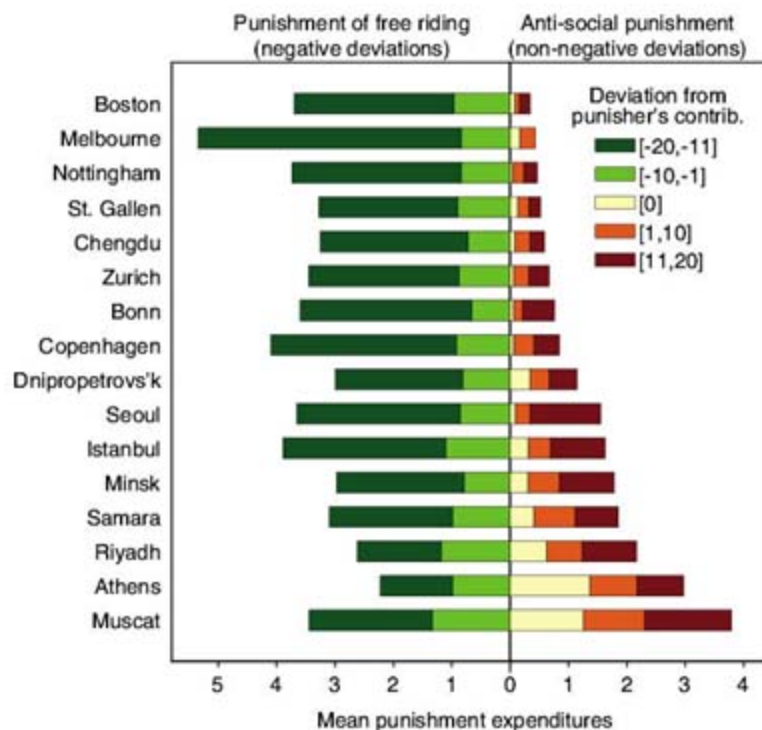
cumstances, some methodological challenges arose. First, with regard to procedures, we followed the rules established in experimental economics (13). A second challenge was maximizing participant pool comparability to avoid confounds of participant pool differences with variations in sociodemographic composition. To minimize sociodemographic variability, we conducted all experiments with university undergraduates ( $n = 1120$ ) who were similar in age, shared an (upper) middle class background, and usually did not know each other. We administered a postexperimental questionnaire to be able to control for further sociodemographic background characteristics (see table S2 for details).

**Results.** We first analyze people's punishment behavior across participant pools. Our perspective is how an individual who has contributed a certain amount to the public good punishes other group members who contributed either less, the same amount, or more than them. Figure 1 therefore displays punishment expenditures as a function of how much the punished individual's contribution deviated from the contribution of the punisher. We label the punishment of negative deviations punishment of free riding because the punished group member rode free on the punisher's contribution. Put differently, from the perspective of the punisher the target member behaved less prosocially than the punisher. In case the target member contributed the same amount or more, he or she behaved at least as prosocially as the punisher. We therefore call the punishment in these cases antisocial punishment.

Punishment behavior differed strongly across participant pools (Fig. 1). This holds in particular for antisocial punishment. A regression analysis of punishment behavior, which controls for the deviation, period effects, and sociodemographic composition, shows that antisocial punishment differed highly and significantly across participant pools [ $\chi^2(14) = 64.9$ ,  $P = 0.000$ ; tables S3 and S4]. Although there was very little antisocial punishment in some participant pools, in others people punished those who contributed the same or more than them as harshly as those who rode free on them. By contrast, punishment of free riding was only weakly significantly different across participant pools [ $\chi^2(14) = 23.1$ ,  $P = 0.059$ ; tables S3 and S4].

The punishment of free riding is likely triggered by negative emotions that arise from a violation of fairness norms and from feeling exploited (1, 2, 18). But what explains antisocial punishment? One plausible reason is that people might not accept punishment and therefore seek revenge (8–10). Revenge is a "human universal" (19) and part of a culture of honor in many societies. Our measure for vengeful punishment is the punishment people mete out as a function of received punishment in the previous period. Controlling for contributions of the punisher and the punished participant, we find a highly significant increase in antisocial punishment across all participant pools as a function of the amount of punishment received

**Fig. 1.** Mean punishment expenditures for a given deviation from the punisher's contribution. The deviations of the punished participant's contribution from the punisher's contribution are grouped into five intervals, where  $[-20, -11]$  indicates that the punished participant contributed between 11 and 20 tokens less than the punishing participant,  $[-10, -1]$  indicates that the punished participant contributed between 1 and 10 tokens less than the punishing participant,  $[0]$  indicates that the punished participant contributed exactly the same amount as the punishing participant,  $[1, 10]$  indicates that the punished participant contributed between 1 and 10 tokens more than the punishing participant, and  $[11, 20]$  indicates that the punished participant contributed between 11 and 20 tokens more than the punishing participant. In Boston, for example, participants (including nonpunishers) expended 0.96 money units on average for all cases of negative deviations between  $[-10, -1]$  and 2.74 money units on average in cases of deviations between  $[-20, -11]$ . Participant pools are sorted according to their mean antisocial punishment. Fig. S2 and tables S3 and S4 provide complementary analyses.





in the previous period. Broken down by participant pools, the effect is highly significant (at  $P < 0.01$ ) in seven participant pools, weakly significantly positive in two participant pools, insignificantly positive in six participant pools, and insignificantly negative in one (tables S3 and S4).

The presence of a punishment opportunity had dramatic consequences on the achieved cooperation levels (Fig. 2). Contributions were highly significantly different across participant pools [Kruskal-Wallis test with group averages over all 10 periods as independent observations,  $\chi^2(15) = 113.1$ ;  $P = 0.000$ ]. Cooperation was stabilized in all participant pools but at vastly different levels (Fig. 2A). Cooperation in about half of the participant pools remained at the initial level (period 1 of the P experiment), whereas contributions increased over time in the others (table S5). The most-cooperative participant pool (in which people contributed 90% of their endowment, on average) contributed 3.1 times as much as the least-cooperative participant pool (with an average contribution of 29% of the endowment). The differences in cooperation across participant pools are significantly negatively related to antisocial punishment: The higher antisocial punishment is in a participant pool, the lower is the average cooperation level in that participant pool (Fig. 2B).

As a consequence of the different patterns of punishment and cooperation, there were also substantial participant-pool differences in earnings in the P experiment. The average per-period earnings differed by more than 250 percentage points between the participant pool with the highest average earnings and that with the lowest average earnings (fig. S3 and table S6).

An important reason for the large participant pool differences in cooperation rates is the fact that participant pools reacted very differently to punishment received. Regression analyses (table S7) show that, in all but one participant pool, people who contributed less than the group average in period  $t$  and who were subsequently punished increased their contribution in period  $t + 1$ . The increase is only significant (at  $P < 0.05$ ) in 11 participant pools, however, and the extent of the mean estimated increase per punishment point received varies considerably between participant pools. Thus, punishment did not have an equally strong disciplinary effect on free riders in all participant pools in the sense of steering low contributors toward higher contributions; in some participant pools, punishment had no cooperation-enhancing effect at all.

The disciplinary effectiveness of punishment for below-average contributions is associated with the extent of antisocial punishment in a participant pool. There is a strong negative correlation between the mean antisocial punishment in a participant pool and the regression coefficient that measures the mean increase per punishment point received for a below-average contribution (Spearman's  $\rho = -0.87$ ,  $P = 0.000$ ,  $n = 16$ ). One explanation is that the prospect of getting punished for at- or above-average contributions in some participant pools limits the low

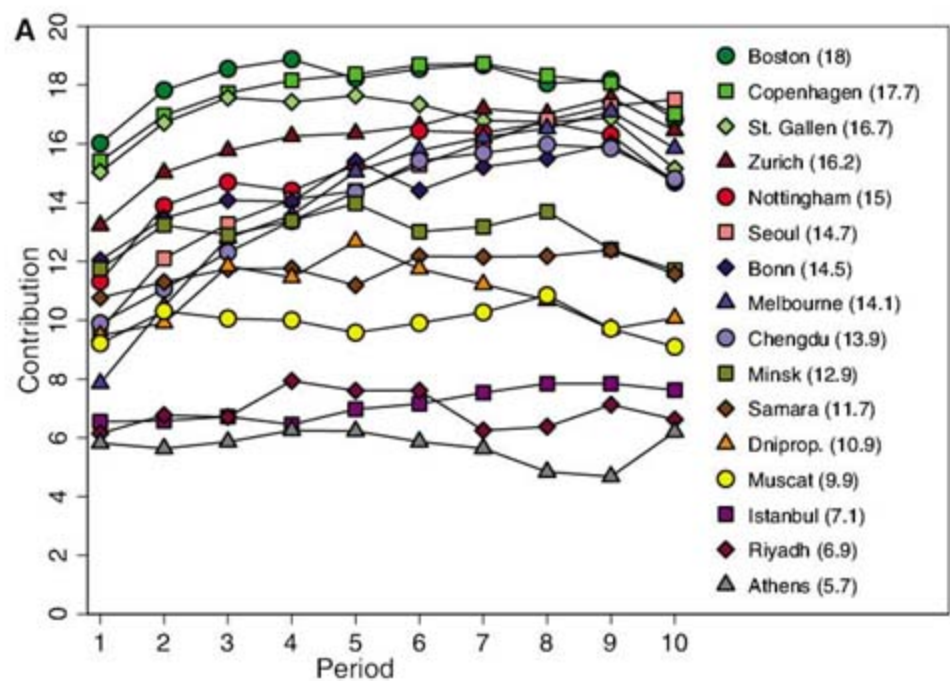
contributors' incentives to increase their contributions. Another explanation has to do with how people perceive the moral message behind punishment because there is evidence that even nonmonetary sanctions (which signal social disapproval) can induce low contributors to increase their contributions (20). Participant pools might have differed in the extent to which people feel ashamed when being punished for low contributions.

A regression analysis (Table 1) summarizes our findings on the impact of punishment on cooperation. To also account for variations of punishment in different groups within participant pools, we use the group average contributions as independent observations.

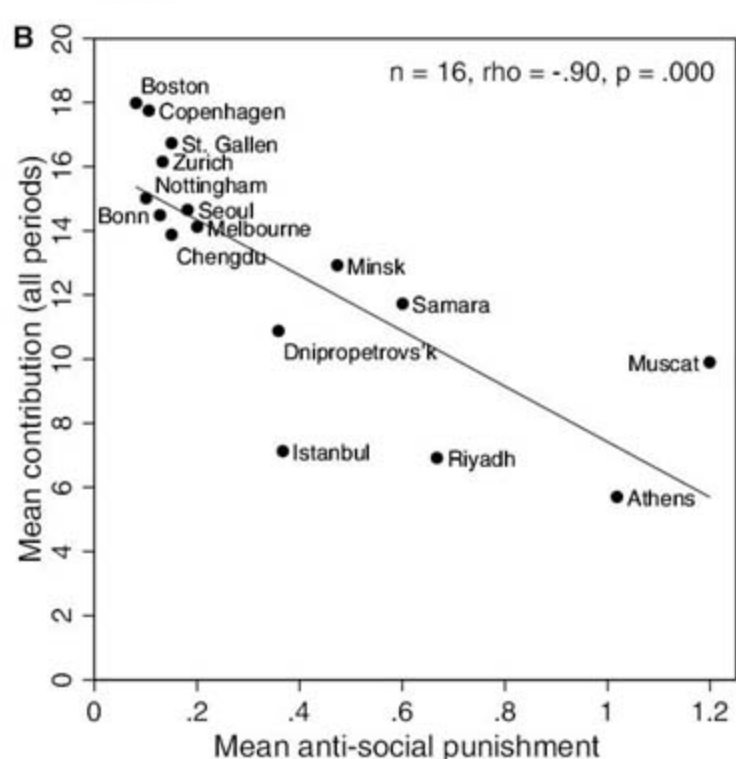
The results show that groups that started at high levels in period 1 of the P experiment also had high group average contributions over the

remaining periods 2 to 10; groups that started at low levels in period 1 of the P experiment had low group average contributions over the remaining periods. Group average punishment of free riding relative to the punishers' own contributions is positively correlated with this group's average contribution, ceteris paribus. The opposite conclusion holds for antisocial punishment.

We also found significant participant pool differences in the N experiment, which serves as a benchmark for the P experiment [Kruskal-Wallis test with group averages over all 10 periods as independent observations,  $\chi^2(15) = 46.5$ ,  $P = 0.000$ ]. Mean contributions varied between 4.9 and 11.5 tokens of the least- and most-cooperative participant pool, respectively (Fig. 3). The span of 6.6 tokens between the least- and most-cooperative participant pool was thus substantially lower than the



**Fig. 2.** (A) Mean contributions to the public good over the 10 periods of the P experiment. Each line corresponds to the average contribution of a particular participant pool. The numbers in parentheses indicate the mean contribution (out of 20) in a particular participant pool. (B) Mean antisocial punishment and mean contribution (across all periods) per participant pool. Rho indicates Spearman rank order correlation between participant pool averages.



span of 12.3 tokens in the P experiment (Fig. 2A). Moreover, in contrast to the P experiment, where contributions were stabilized at vastly different levels, contributions in the N experiment dwindled to lower levels almost everywhere (table S8 and fig. S4).

Compared with the N experiment, the presence of a punishment option had at least a weakly significant cooperation-enhancing effect in 11

participant pools (Wilcoxon signed ranks test with independent group average contribution rates across all periods as observations, fig. S4 and table S9); the change in cooperation between the N and the P experiment was not significant in the other five participant pools. Thus, the cooperation-enhancing effect of a punishment opportunity cannot be taken for granted. This

finding stands in contrast to previous results from experiments conducted in the United States and Western Europe, where punishment always increased cooperation in experiments with comparable fixed-group designs and parameters (8, 10, 20–22).

The reason for this result is related to antisocial punishment: the higher antisocial punishment was in a participant pool, the lower was the rate of increase in cooperation in the P experiment relative to the N experiment (Spearman's  $\rho = -0.76$ ,  $P = 0.001$ ,  $n = 16$ ). Furthermore, participant pools' average cooperation levels in period 1 of the P experiment (where participants had not yet acquired any experience with punishment) were significantly negatively correlated with their subsequent mean expenditures on antisocial punishment: The more a participant pool expended on antisocial punishment in the later stages of the experiment, the lower was its initial cooperation level (Spearman's  $\rho = -0.78$ ,  $P = 0.000$ ,  $n = 16$ ).

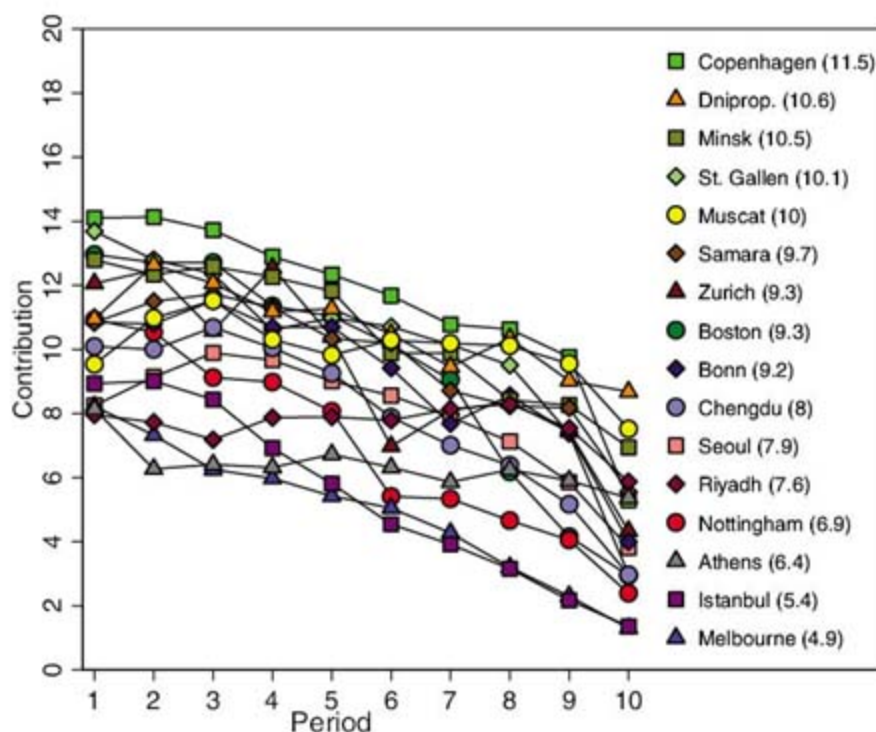
What explains the large participant pool differences in antisocial punishment and hence cooperation levels? Punishment may be related to social norms of cooperation. Social norms exist at a macrosocial level and refer to widely shared views about acceptable behaviors and the deviations subject to possible punishment (23, 24). Thus, if participant pools held different social norms with regard to cooperation and free riding, they actually might have punished differently. An interesting set of relevant social norms are norms of civic cooperation (14) as they are expressed in people's attitudes to tax evasion, abuse of the welfare state, or dodging fares on public transport. These are all situations that can be modeled as public goods problems. The stronger norms of civic cooperation are in a society, the more free riding might be viewed as unacceptable and the more it might be punished in consequence. The flip side of the argument is that cooperators, who behave in the normatively desirable way, should not get punished; strong norms of civic cooperation might act as a constraint on antisocial punishment.

The strength of the rule of law in a society might also have an impact on antisocial punishment. If the rule of law is strong, people trust the law enforcement institutions, which are perceived as being effective, fair, impartial, and bound by the law (25). Revenge is shunned. If the rule of law is weak, the opposite holds. Thus, the rule of law reflects how norms are commonly enforced in a society.

We construct the variable norms of civic cooperation from data taken from the World Values Survey (13) (fig. S1A). The variable is derived from answers of a large number of selected representative residents of a country to questions on how justified (on a 10-point scale; 1 is fully justified; 10 is never justified) people think tax evasion, benefit fraud, or dodging fares on public transport are. The more reproachful these behaviors are in the eyes of the average citizen, the stronger are a society's norms of civic coopera-

**Table 1.** Punishment and cooperation levels. Ordinary least squares regressions with the group average contributions of all groups, which show any variation in contributions as independent observation ( $n = 273$ ). The group average contributions over periods 2 to 10 are the dependent variables. The independent variables are the group average contributions in period 1, the group averages of punishment points assigned to group members who contributed less than the punishing participant (group average punishment of free riding) and to group members who are equally or more cooperative than the punishing participant (group average antisocial punishment). Model 1 does not control for the mean cooperation level in a participant pool, whereas model 2 controls for it by adding participant pool dummies. The adjusted  $r^2$  increases by only 7% and the results remain robust, although the coefficient for antisocial punishment is reduced in size. \* $P < 0.1$ , \*\* $P < 0.05$ , \*\*\* $P < 0.01$ . Numbers in parentheses indicate robust standard errors.

	Dependent variable: group average contributions in period 2 to 10	
	1	2
Group average contributions in period 1	0.779*** (0.052)	0.720*** (0.065)
Group average punishment of free riding	0.521** (0.201)	0.480** (0.200)
Group average antisocial punishment	-2.247*** (0.350)	-1.256*** (0.325)
Constant	5.057*** (0.688)	5.899*** (1.221)
Participant pool dummies	No	Yes
Adjusted $r^2$	0.60	0.67
F test	136.9	31.3
P value	0.000	0.000
N	273	273



**Fig. 3.** Mean contributions to the public good over the 10 periods of the N experiment. Each line corresponds to the average contribution of a particular participant pool. The numbers in parentheses indicate the mean contribution (out of 20) in a particular participant pool.

tion (14). The country scores of our 16 participant pools vary between 6.91 and 9.79 (the mean is 8.53); the available world sample range ( $n = 81$  countries; mean = 8.64) lies between 6.75 and 9.81. Thus, the societies of our participant pools cover almost the whole available worldwide range of the distribution of norms of civic cooperation.

The rule of law indicator (13) (fig. S1B) is based on a host of different variables that measure “the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, the police, and the courts, as well as the likelihood of crime and violence” (25). The theoretical range is  $-2.5$  (very weak rule of law) to  $2.5$  (very strong rule of law). The empirically observed range of the 211 countries for which this indicator is available is  $-2.20$  to  $1.99$ . The rule of law indicator varies between  $-1.23$  and  $1.96$  in the countries of our participant pools.

Because both indicators reflect the views of the average citizen in a given society, it is likely that our participants, through various forms of cultural transmission (26), have been exposed to the prevalent social norms and have perceptions of the quality of the rule of law in their respective societies. Moreover, previous research, conducted in small-scale societies, suggests that experimentally observed behavior reflects socioeconomic conditions and daily experiences (11). Thus, there are good reasons to expect that the experimentally observable punishment behavior might be correlated with our indicators.

We investigated the link between punishment and the two indicators econometrically by running regressions of punishment expenditures on the variables norms of civic cooperation and rule of law. We distinguished between punishment of

free riding and antisocial punishment, and we also controlled for the punisher’s contribution, the contribution of the punished participant, the contribution of other group members, period effects, and the individual socioeconomic characteristics (to control for differences in participant-pool composition). The estimation method is Tobit (with robust standard errors clustered on the independent group).

The estimation results (Table 2) show that the stronger norms of civic cooperation are in the society, the harder people in the respective participant pool punish those who contributed less than them ( $P < 0.01$ ). Rule of law has an insignificantly positive impact on the punishment of free riding. With respect to antisocial punishment, we found that both norms of civic cooperation and rule of law are significantly negatively correlated with punishment (at  $P < 0.05$ ). In other words, antisocial punishment is harsher in participant pools from societies with weak norms of civic cooperation and a weak rule of law. Additional analyses (table S10) show that antisocial punishment also varies highly significantly with a variety of indicators developed by social scientists in order to characterize societies (table S1). Thus, the extent of antisocial punishment is most likely affected by the wider societal background.

**Discussion.** Evidence from economics, sociology, political science, and anthropology suggests that human social groups differ strongly in how successfully they solve cooperation problems (14, 27–29). In reality, many exogenous factors, institutional and environmental conditions as well as population characteristics, can explain varying degrees of cooperative success. Our contribution is to show experimentally that (antisocial) punishment can lead to very strong differences

in cooperation levels among comparable social groups acting in identical environments.

Antisocial punishment of cooperators existed in all our participant pools, but its importance and detrimental consequences varied strongly across them. Revenge is a likely explanation for antisocial punishment in most participant pools, but other (population-specific) motives might be relevant as well. Some antisocial punishment may be efficiency-enhancing in intent to induce the punished individual to increase his or her contributions. The fact that in most participant pools antisocial punishment was lower the higher the punished participant’s contribution was is consistent with this explanation (table S7A). Because punishment in our experiment was cheaper for the punisher than for the punished participant, people with a strong taste for dominance (30), a competitive personality (31), or a desire to maximize relative payoffs (32) might not only punish freeloaders but also cooperators, even including those who contributed the same amount as the punisher. Low contributors might also view high contributors as do-gooders who have shown them up. Punishment may therefore be an act of “do-gooder derogation” (33). Similarly, as observed in some bargaining experiments (12, 34, 35) in which people reject hyperfair offers, people for various reasons might be suspicious of others who appear too generous. Normative conformity, a desire and expectation to behave as all others do, is part of human psychology (36) and may lead to the punishment of all deviators, cooperators, and free riders alike. Punishment may be also related to in-group-out-group distinctions (37) because people might retaliate if punished by an out-group member (38). Societies also differ in the extent to which their social structures are governed by in-group-out-group distinctions. For instance, according to some cross-cultural psychologists (15, 39) in “collectivist” societies many interactions are confined to close-knit social networks, whereas in “individualistic” societies interactions are more permeable across social groups. Because in our experiment all participants were strangers to one another, people in collectivist societies might be more inclined than people in individualistic societies to perceive other participants as out-group members. Therefore, antisocial punishment might be stronger in collectivist than in individualistic societies. Our evidence is consistent with this possibility because in regressions similar to those of Table 2 antisocial punishment is highly significantly correlated with a widely used societal-level measure of individualism-collectivism (15) (table S10).

Our finding that social norms of cooperation and punishment are linked is of relevance for the debate about social capital (14) and in particular a literature that argues that informal sanctions often substitute for formal enforcement mechanisms if these are lacking or not working well (7, 27, 40–42). The fact that antisocial punishment is negatively correlated

**Table 2.** Punishment, norms of civic cooperation, and the rule of law. The dependent variable is assigned punishment points to participants who contributed less than the punishing participants (models 1 to 3) or to participants who contributed the same or more than the punishing participant (models 4 to 6). The independent variables are the country scores of norms of civic cooperation and rule of law. Controls include the participants’ own contribution, the punished participant’s contribution, the average contribution of the remaining two participants, the period, a dummy for the final period, and individual socioeconomic characteristics. We show the coefficients of Tobit estimates (43). Robust standard errors are calculated by using the group as the independent cluster. Table S10 contains further analyses.

	Punishment of free riding (negative deviations)			Antisocial punishment (nonnegative deviations)		
	1	2	3	4	5	6
Norms of civic cooperation	0.410*** (0.154)		0.371** (0.155)	-1.093*** (0.326)		-0.740** (0.371)
Rule of law		0.164 (0.111)	0.067 (0.110)		-0.641*** (0.221)	-0.618** (0.254)
Constant	-5.047*** (1.400)	-1.843*** (0.469)	-4.708*** (1.398)	5.622* (2.900)	-3.479*** (0.719)	2.422 (3.360)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
$\sigma$	3.323	3.457	3.322	5.583	5.665	5.566
$P$ value	0.000	0.000	0.000	0.000	0.000	0.000
log pseudolikelihood	-12203	-13299	-12202	-10574	-11989	-10539
$N$	8350	8950	8350	19850	20660	19850

with the strength of the rule of law and also with cooperation levels suggests that the quality of the formal law enforcement institutions and informal sanctions are complements (rather than substitutes). Informal sanctions might be more effective in sustaining voluntary cooperation when the formal law enforcement institutions operate more effectively because antisocial punishment is lower in these societies. The detrimental effects of antisocial punishment on cooperation (and efficiency) also provide a further rationale why modern societies shun revenge and centralize punishment in the hands of the state.

#### References and Notes

- E. Fehr, S. Gächter, *Nature* **415**, 137 (2002).
- D. J. F. de Quervain *et al.*, *Science* **305**, 1254 (2004).
- Ö. Gülerk, B. Irlenbusch, B. Rockenbach, *Science* **312**, 108 (2006).
- B. Rockenbach, M. Milinski, *Nature* **444**, 718 (2006).
- R. Axelrod, W. Hamilton, *Science* **211**, 1390 (1981).
- M. A. Nowak, K. Sigmund, *Nature* **437**, 1291 (2005).
- E. Ostrom, J. M. Walker, R. Gardner, *Am. Polit. Sci. Rev.* **86**, 404 (1992).
- N. Nikiforakis, *J. Public Econ.* **92**, 91 (2008).
- M. Cinyabuguma, T. Page, L. Putterman, *Exp. Econ.* **9**, 265 (2006).
- L. Denant-Boemont, D. Masclet, C. N. Noussair, *Econ. Theory* **33**, 145 (2007).
- J. Henrich *et al.*, *Behav. Brain Sci.* **28**, 795 (2005).
- J. Henrich *et al.*, *Science* **312**, 1767 (2006).
- Materials and methods are available on Science Online.
- S. Knack, P. Keefer, *Q. J. Econ.* **112**, 1251 (1997).
- G. Hofstede, *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations* (Sage, Thousand Oaks, CA, 2001).
- R. Inglehart, W. E. Baker, *Am. Sociol. Rev.* **65**, 19 (2000).
- U. Fischbacher, *Exp. Econ.* **10**, 171 (2007).
- A. Falk, E. Fehr, U. Fischbacher, *Econometrica* **73**, 2017 (2005).
- J. Elster, *Ethics* **100**, 862 (1990).
- D. Masclet, C. Noussair, S. Tucker, M. C. Villeval, *Am. Econ. Rev.* **93**, 366 (2003).
- E. Fehr, S. Gächter, *Am. Econ. Rev.* **90**, 980 (2000).
- T. Page, L. Putterman, B. Unel, *Econ. J.* **115**, 1032 (2005).
- J. Coleman, *Foundations of Social Theory* (Belknap, Cambridge, MA, 1990).
- J. Henrich, N. Henrich, *Why Humans Cooperate: A Cultural and Evolutionary Explanation*, Evolution and Cognition Series (Oxford Univ. Press, Oxford, 2007).
- D. Kaufmann, A. Kraay, M. Mastruzzi, World Bank Policy Research Working Paper No. 4280, [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=999979](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=999979)
- R. Boyd, P. J. Richerson, *Culture and the Evolutionary Process* (Univ. of Chicago Press, Chicago, 1985).
- E. Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action, the Political Economy of Institutions and Decisions* (Cambridge Univ. Press, Cambridge, 1990).
- R. J. Sampson, S. W. Raudenbush, F. Earls, *Science* **277**, 918 (1997).
- R. B. Edgerton, *Sick Societies: Challenging the Myth of Primitive Harmony* (Free Press, New York, 1992).
- T. H. Clutton-Brock, G. A. Parker, *Nature* **373**, 209 (1995).
- W. B. G. Liebrand, R. W. T. L. Jansen, V. M. Rijken, C. J. M. Suhre, *J. Exp. Soc. Psychol.* **22**, 203 (1986).
- K. Fliessbach *et al.*, *Science* **318**, 1305 (2007).
- B. Monin, *Int. Rev. Soc. Psychol.* **20**, 53 (2007).
- D. L. Bahry, R. K. Wilson, *J. Econ. Behav. Organ.* **60**, 37 (2006).
- H. Hennig-Schmidt, Z.-Y. Li, C. Yang, *J. Econ. Behav. Organ.* **65**, 373 (2008).
- J. Henrich, *J. Econ. Behav. Organ.* **53**, 3 (2004).
- H. Bernhard, U. Fischbacher, E. Fehr, *Nature* **442**, 912 (2006).
- J.-K. Choi, S. Bowles, *Science* **318**, 636 (2007).
- H. C. Triandis, *Individualism and Collectivism*, R. E. Nisbett, Ed., *New Directions in Social Psychology* (Westview, Boulder, CO, 1995).
- A. Dixit, *Lawlessness and Economics: Alternative Modes of Economic Governance* (Princeton Univ. Press, Princeton, NJ, 2004).
- E. A. Posner, *Law and Social Norms* (Harvard Univ. Press, Cambridge, MA, 2000).
- R. C. Ellickson, *Order Without Law: How Neighbors Settle Disputes* (Harvard Univ. Press, Cambridge, MA, 1991).
- Alternative estimation methods, like Probit and Poisson, yield very similar results, both in terms of signs and statistical significance.
- All authors contributed equally to this work. The authors thank various workshop audiences, in particular the Arts and Humanities Research Council workshops Culture and the Mind in Sheffield, and I. Bohnet, R. Boyd, S. Burks, E. Fehr, U. Fischbacher, D. Gambetta, H. Gintis, G. Grimalda, J. Henrich, P. Richerson, B. Rockenbach, R. Sapolsky, and R. Zeckhauser for helpful discussions. We are grateful for financial support from the University of Nottingham, the Grundlagenforschungsfonds at the University of St. Gallen, the Latsis Foundation (Geneva), and the EU-TMR Research Network ENDEAR (FMRX-CT98-0238). This paper is part of the MacArthur Foundation Network on Economic Environments and the Evolution of Individual Preferences and Social Norms.

#### Supporting Online Material

[www.sciencemag.org/cgi/content/full/319/5868/1362/DC1](http://www.sciencemag.org/cgi/content/full/319/5868/1362/DC1)

Materials and Methods

SOM Text

Figs. S1 to S4

Tables S1 to S10

References and Notes

5 December 2007; accepted 28 January 2008

10.1126/science.1153808

## REPORTS

# Fiber-Optical Analog of the Event Horizon

Thomas G. Philbin,<sup>1,2</sup> Chris Kuklewicz,<sup>1</sup> Scott Robertson,<sup>1</sup> Stephen Hill,<sup>1</sup> Friedrich König,<sup>1</sup> Ulf Leonhardt<sup>1\*</sup>

The physics at the event horizon resembles the behavior of waves in moving media. Horizons are formed where the local speed of the medium exceeds the wave velocity. We used ultrashort pulses in microstructured optical fibers to demonstrate the formation of an artificial event horizon in optics. We observed a classical optical effect: the blue-shifting of light at a white-hole horizon. We also showed by theoretical calculations that such a system is capable of probing the quantum effects of horizons, in particular Hawking radiation.

Laboratory analogs of black holes (1–3) are inspired by a simple and intuitive idea (4): The space-time geometry of a black hole resembles a river (5, 6)—a moving medium flowing toward a waterfall, the singularity. Imagine

that the river carries waves propagating against the current with speed  $c'$ . The waves play the role of light, where  $c'$  represents  $c$ , the speed of light in vacuum. Suppose that the closer the river gets to the waterfall, the faster it flows, and that at some point the speed of the river exceeds  $c'$ . Clearly, beyond this point waves can no longer propagate upstream. The point of no return corresponds to the horizon of the black hole. Imagine another situation: a fast river flowing out into the sea, getting slower. Waves cannot enter the river beyond the point where the flow

speed exceeds the wave velocity; the river resembles an object that nothing can enter: a white hole.

Nothing, not even light, can escape from a gravitational black hole. Yet according to quantum physics, the black hole is not entirely black but emits waves that are in thermal equilibrium (7–9). The waves consist of correlated pairs of quanta; one originates from inside and the other from outside the horizon. Seen from one side of the horizon, the gravitational black hole acts as a thermal black-body radiator sending out Hawking radiation (7–9). The effective temperature depends on the surface gravity (7–9), which, in the analog model, corresponds to the flow-velocity gradient at the horizon (1–5).

The Hawking temperature of typical black holes lies far below the temperature of the cosmic microwave background, so an observation of Hawking radiation in astrophysics seems unlikely. However, laboratory demonstrations of analogs of Hawking radiation could be feasible. One type of recent proposal (10–12) suggests the use of ultracold quantum gases such as alkali Bose-Einstein condensates or ultracold alkali fermions (12). When a condensate in a wave guide is pushed over a potential barrier, it may exceed the speed of sound (typically a few millimeters per second)

<sup>1</sup>School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews, Fife, KY16 9SS, UK. <sup>2</sup>Max Planck Research Group of Optics, Information and Photonics, Günther-Scharowsky-Strasse 1, Bau 24, D-91058 Erlangen, Germany.

\*To whom correspondence should be addressed. E-mail: [ulf@st-andrews.ac.uk](mailto:ulf@st-andrews.ac.uk)

and is calculated to generate a Hawking temperature of about 10 nK (11).  $^3\text{H}$  offers a multitude of analogs between quantum fluids and the standard model, including Einsteinian gravity (2). For example, the analogy between gravity and surface waves in fluids (13) has inspired ideas for artificial event horizons at the interface between two sliding superfluid phases (14), but so far, none of the quantum features of horizons has been measured in  $^3\text{H}$ . Proposals for optical black holes (15, 16) have relied on slowing down light (17) so that it matches the speed of the medium (15) or on bringing light to a complete standstill (16), but in these cases absorption may pose a severe problem near the horizon, where the spectral transparency window (17) vanishes.

On the other hand, is it necessary to physically move a medium to establish a horizon? What really matters are only the effective properties of the medium. If they change, for example as a propagating front, but the medium itself remains at rest, the situation is essentially indistinguishable from that in a moving medium. Such ideas have been discussed for moving solitons and domain walls (18) in superfluid  $^3\text{H}$  (2) and more recently for microwave transmission lines with variable capacity (19), but they have remained impractical so far.

Here we report the first experimental observation of the classical optical effects of horizons, the blue-shifting of light at a white-hole horizon, and show theoretically that our scheme combines several promising features for demonstrating quantum Hawking radiation in the optical domain. Our idea, illustrated in Fig. 1, is based on the nonlinear optics of ultrashort light pulses in optical fibers (20), where we exploit the remarkable control of nonlinearity, birefringence, and dispersion that is possible in microstructured fibers (21, 22).

Using a Ti:sapphire laser, we created 70-fs-long nondispersive pulses (solitons) at a carrier wavelength of 803 nm and a repetition rate of 80 MHz inside 1.5 m of microstructured fiber (NL-PM-750B from Crystal Fiber, Birkerød, Denmark). Each pulse modifies the optical properties of the fiber because of the Kerr effect (20): The effective refractive index of the fiber,  $n_0$ , gains an additional contribution  $\delta n$  that is proportional to the instantaneous pulse intensity  $I$  at position  $z$  and time  $t$

$$n = n_0 + \delta n, \quad \delta n \propto I(z, t) \quad (1)$$

This contribution to the refractive index  $n$  moves with the pulse. The pulse thus establishes a moving medium, although nothing material is moving. This effective medium naturally moves at the speed of light in the fiber, because it is made by light itself.

We also launched a continuous wave of light, a probe, that follows the pulse with slightly higher group velocity, attempting to overtake it. In order to distinguish the probe from the pulse, it

oscillates at a significantly different frequency  $\omega$ . Our probe-light laser is tunable over wavelengths  $2\pi c/\omega$  from 1460 to 1540 nm. While approaching the pulse, the Kerr contribution  $\delta n$  slows down the probe until the probe's group velocity reaches the speed of the pulse. The trailing end of the pulse establishes a white-hole horizon, an object that light cannot enter, unless it tunnels through the pulse. Conversely, the front end creates a black-hole horizon for probe light that is slower than the pulse. Because  $\delta n$  is small, the initial group velocity of the probe should be sufficiently close to the speed of the pulse. In microstructured fibers (22), the group-velocity dispersion of light is engineered by arrangements of air holes (submicrometer-wide hollow cylinders along the fiber). We selected a fiber in which the group velocity of pulses near the 800-nm carrier wavelength of mode-locked Ti:sapphire lasers matched the group velocities of probes in the infrared telecommunication band around 1500 nm; standard optical fibers (20) do not have this property.

At the horizon of an astrophysical black hole, light freezes, reaching wavelengths shorter than the Planck scale, where the physics is unknown. [The Planck length is given by  $\sqrt{\{hG/c^3\}}$ , where  $G$  is the gravitational constant.] Some elusive trans-Planckian mechanism must regularize this behavior (23, 24). In our case, the fiber-optical analog of trans-Planckian physics is known and simple—it is contained in the frequency dependence of the refractive index  $n$ , the dispersion of the fiber. At the trailing end of the pulse, the incoming probe modes are compressed, oscillating with increasing frequency; they are blue-shifted. In turn, the dispersion limits the frequency shifting by tuning the probe out of the horizon. In the case of normal group-velocity dispersion, the blue-shifted light falls behind. At the black-hole horizon, the reverse occurs: A probe slower than the pulse is red-shifted and then moves ahead of the pulse.

**Fig. 1. Fiber-optical horizons.** (A) A light pulse in a fiber slows down infrared probe light, attempting to overtake it. The diagrams below are in the co-moving frame of the pulse. (B) Classical horizons. The probe is slowed down by the pulse until its group velocity matches the pulse speed at the points indicated by black dots, establishing a white-hole horizon at the back and a black-hole horizon at the front of the pulse. The probe light is blue-shifted at the white hole until the optical dispersion releases it from the horizon. (C) Quantum pairs. Even if no probe light is incident, the horizon emits photon pairs corresponding to waves of positive frequencies from the outside of the horizon paired with waves at negative frequencies from beyond the horizon. An optical shock has steepened the pulse edge, increasing the luminosity of the white hole.

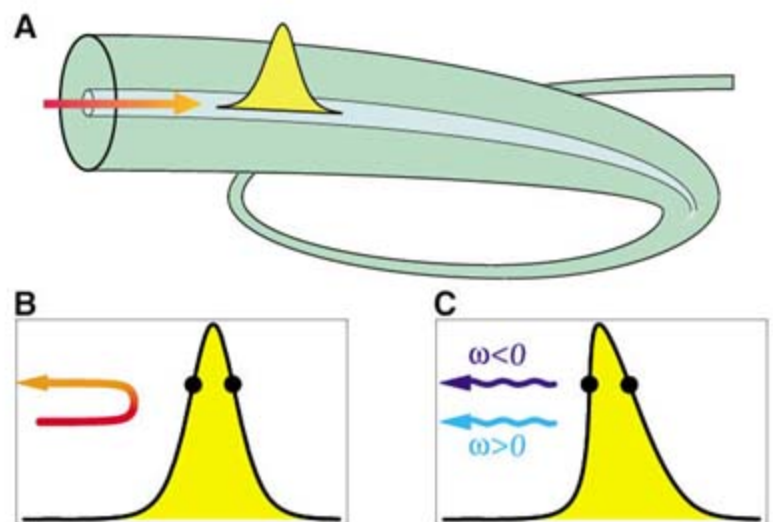


Figure 2 shows the difference in the spectrum of the probe light (incident with  $\omega_1$ ) with and without the pulses, clearly displaying a blue-shifted peak at  $\omega_2$ . To quantitatively describe this effect, we consider the frequency  $\omega'$  in a co-moving frame that is connected to the laboratory frame frequency  $\omega$  by the Doppler formula

$$\omega' = \left(1 - \frac{vu}{c}\right)\omega \quad (2)$$

For a stable pulse,  $\omega'$  is a conserved quantity, whereas  $\omega$  follows the contours of fixed  $\omega'$  when  $\delta n$  varies with the intensity profile of the pulse (Fig. 3). For sufficiently large  $\delta n$ , the frequency  $\omega$  completes an arch from the initial  $\omega_1$  to the final  $\omega_2$ ; it is blue-shifted by the white-hole horizon. At a black-hole horizon, the arch is traced the other way round from  $\omega_2$  to  $\omega_1$ . For the frequency at the center of the arches, an infinitesimal  $\delta n$  is sufficient to cause a frequency shift; at this frequency, the group velocity of the probe matches the group velocity of the pulse. Figure 2 shows that both the blue-shifted and probe light are spectrally broadened. These features are easily explained: The horizon acts only during the time while probe and pulse propagate in the fiber, where only a finite fraction of the probe is frequency-shifted, forming a blue-shifted pulse and also a gap in the probe light, a negative pulse; these pulses have a characteristic spectrum with a width that is inversely proportional to the fiber length. We compared the measured spectra with the theory of light propagation in the presence of horizons and found very good agreement (25).

Imagine instead of a single probe a set of probe modes. The modes should be sufficiently weakly excited that they do not interact with each other via the Kerr effect, but they experience the cross Kerr effect of the pulse, the presence of the medium (Eq. 1) moving with the velocity  $u$ . The modes constitute a quantum field of light in a moving medium (25, 26). Classical light is a real electro-

magnetic wave. So, according to Fourier analysis, any amplitude oscillating as  $\exp(-i\omega t)$  at a positive angular frequency  $\omega$  must be accompanied by the complex conjugate amplitude at  $-\omega$ . In quantum field theory (8, 9, 25), the positive-frequency modes correspond to the annihilation operators and the negative-frequency modes to the creation operators (26). Processes that mix positive and negative frequencies in the laboratory frame (in the glass of the fiber) thus create observable light quanta.

In the near ultraviolet around 300 nm, the dispersion of microstructured fibers (22) is dominated by the bare dispersion of glass, where  $n_0$  rapidly grows with frequency (20) exceeding  $c/u$ . For such ultraviolet modes, the medium moves at superluminal speed. We see from the Doppler formula (Eq. 2) that these superluminal modes oscillate with negative frequencies  $\omega'$  in the co-moving frame for positive frequencies  $\omega$  in the laboratory frame, and vice versa. Moreover, each subluminal mode with positive  $\omega$  has a superluminal partner oscillating at the same co-moving frequency  $\omega'$  but with negative laboratory frequency. The pulse does not change  $\omega'$ , but it may partially convert sub- and superluminal partner

modes into each other, thus creating photons (8, 9). Even if all the modes are initially in their vacuum states, the horizon spontaneously creates photon pairs. This process represents the optical analog of Hawking radiation (7), and it occurs at both the black-hole and white-hole horizon (25). Photons with positive  $\omega'$  correspond to the particles created at the outside of the black hole (8, 9), whereas the negative-frequency photons represent their partners beyond the horizon. In our case, the photon pairs are distinguishable from the intense pulse, because their polarization can be orthogonal to the pulse and their frequencies differ from the carrier frequency by an octave. Furthermore, one can discriminate the Hawking effect from other nonlinear optical processes, such as four-wave mixing, because it is not subject to their phase-matching conditions (20). Moreover, in addition to observing Hawking radiation per se, one could detect the correlations of the Hawking partners—a feat that is impossible in astrophysics, because there the partner particles are lost beyond the horizon of the black hole.

In order to give a quantitative argument for the Hawking effect in optical fibers, we describe the propagation distance  $z$  in terms of the time  $\zeta$  it

takes for the pulse to traverse it,  $\zeta = z/u$ , and introduce the retarded time  $\tau = t - z/u$ . The phase  $\varphi$  of each mode evolves as

$$\varphi = -\int(\omega d\tau + \omega' d\zeta) \quad (3)$$

We assume that the mode conversion occurs in a narrow interval of retarded time  $\tau$  near a horizon around  $\tau = 0$ , where we linearize  $\delta n$  in  $\tau$  such that

$$1 - \frac{nu}{c} = \alpha'\tau \quad (4)$$

We obtain from the phase integral (Eq. 3) and the Doppler formula (Eq. 2) the characteristic logarithmic phase at a horizon (8, 9). We use the standard result (8, 9, 25): Hawking radiation is Planck-distributed with the temperature

$$k_B T' = \frac{h\alpha'}{2\pi} \quad (5)$$

where  $k_B$  denotes Boltzmann's constant. For evaluating  $\alpha'$ , we consider  $\delta n$  at  $\tau = 0$ , where

$$\alpha' = -\frac{u}{c} \frac{\partial n}{\partial \tau} \Big|_0 = -\frac{u}{c} \frac{\partial \delta n}{\partial \tau} \Big|_0 \quad (6)$$

Note that  $T'$  denotes the Hawking temperature in the co-moving frame, defined with respect to the Doppler-shifted frequencies  $\omega'$ , a temperature that is characterized by the Doppler-shifted Hawking frequency  $\alpha'$  in regions away from the pulse. We use the Doppler formula (Eq. 2) with the refractive index (Eq. 1) and the linearization (Eq. 4) taken at  $\tau = 0$ , and obtain

$$\alpha' = \left(1 - n_0 \frac{u}{c}\right) \alpha = \frac{u}{c} \delta n \Big|_0 \alpha \quad (7)$$

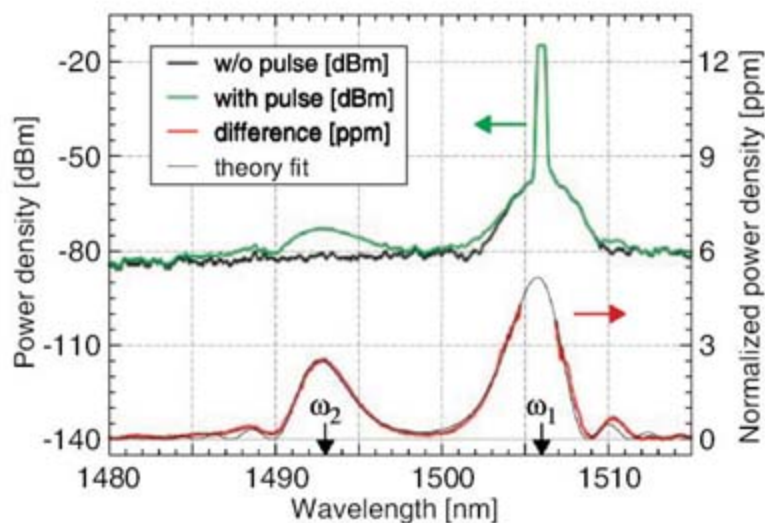
Consequently, the Hawking temperature  $T$  in the laboratory frame is

$$k_B T = \frac{h\alpha}{2\pi}, \quad \alpha = -\frac{1}{\delta n} \frac{\partial \delta n}{\partial \tau} \Big|_0 \quad (8)$$

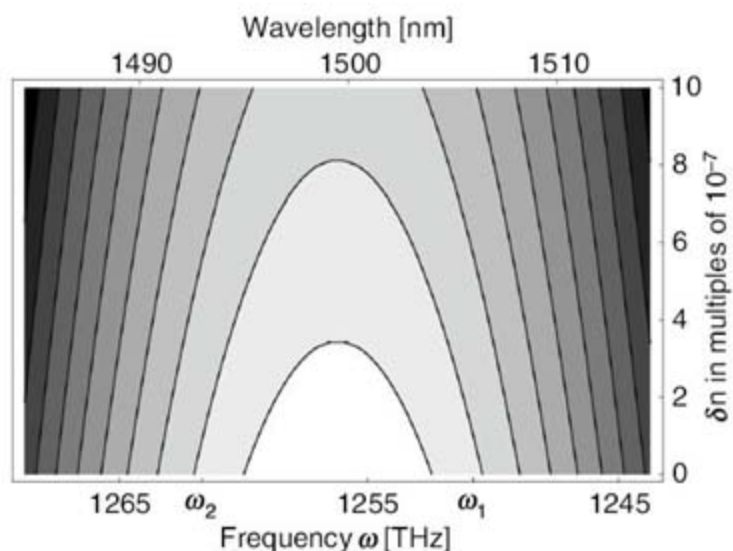
Because  $T$  does not depend on the magnitude of  $\delta n$ , even the typically very small refractive index variations of nonlinear fiber optics (20) may lead to a substantial Hawking temperature when  $\delta n$  varies on the scale of an optical wavelength. This is achievable with optical pulses of a few cycles (27, 28).

The Kerr nonlinearity (20) influences not only the probe modes but the pulse as well (20). This self-Kerr effect shapes the pulse while it propagates in the fiber. Regions of high intensity lag behind, because for them the effective refractive index is increased. The black-hole horizon at the front is stretched, but the trailing edge becomes extremely steep, infinitely steep in theory (20): The pulse develops an optical shock (20). The steep white-hole horizon will dominate the Hawking effect of the pulse. In practice, dispersion combined with other nonlinear optical

**Fig. 2.** Measurement of blue-shifting at a white-hole horizon. The black curve shows the power spectrum of probe light that has not interacted with the pulses, whereas the green curve displays the result of the interaction; both curves are represented on a logarithmic scale. The difference between the spectra on a linear scale, shown in red, exhibits a characteristic peak around the blue-shifted wavelength ( $\omega_2$ ) and another peak around the spectral line of the probe laser ( $\omega_1$ ) due to a gap in the probe light; both features indicate the presence of a horizon. ppm, parts per million.



**Fig. 3.** Doppler contours. The Doppler-shifted frequency  $\omega'$  of the probe is a conserved quantity. The pulse shifts the laboratory frequency  $\omega$  along the contour lines of  $\omega'$  as a function of the instantaneous  $\delta n$ ; the same applies to the wavelength  $\lambda = 2\pi c/\omega$ . If the pulse is sufficiently intense that  $\delta n$  reaches the top of a contour, the probe light completes an arch on the diagram while leaving the pulse; it is red- or blue-shifted, depending on its initial frequency.



processes in the fiber, in particular stimulated Raman scattering (20), limit the optical shock. Assuming that the steepness at the shock front is comparable to twice the frequency of the pulse carrier,  $8 \times 10^{14}$  Hz, the Hawking temperature (Eq. 8) reaches  $10^3$  K, which is many orders of magnitude higher than that in condensed-matter analogs of the event horizon (10–12, 18).

Our scheme thus solves two problems at once in a natural way: how to let an effective medium move at superluminal speed, and how to generate a steep velocity profile at the horizon. Here the various aspects of the physics of artificial black holes conspire together, in contrast to most other proposals (1–4, 10–16).

#### References and Notes

1. *Artificial Black Holes*, M. Novello, M. Visser, G. E. Volovik, Eds. (World Scientific, Singapore, 2002).
2. G. E. Volovik, *The Universe in a Helium Droplet* (Clarendon Press, Oxford, 2003).
3. W. G. Unruh, R. Schützhold, *Quantum Analogues: From Phase Transitions to Black Holes and Cosmology* (Springer, Berlin, 2007).

4. W. G. Unruh, *Phys. Rev. Lett.* **46**, 1351 (1981).
5. T. Jacobson, *Prog. Theor. Phys.* **136** (suppl.), 1 (1999).
6. G. Rousseaux, C. Mathis, P. Maïssa, T. G. Philbin, U. Leonhardt, <http://arxiv.org/abs/0711.4767>.
7. S. M. Hawking, *Nature* **248**, 30 (1974).
8. N. D. Birrell, P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge Univ. Press, Cambridge, 1984).
9. R. Brout, S. Massar, R. Parentani, Ph. Spindel, *Phys. Rep.* **260**, 329 (1995).
10. L. J. Garay, J. R. Anglin, J. I. Cirac, P. Zoller, *Phys. Rev. Lett.* **85**, 4643 (2000).
11. S. Giovanazzi, C. Farrell, T. Kiss, U. Leonhardt, *Phys. Rev. A* **70**, 063602 (2004).
12. S. Giovanazzi, *Phys. Rev. Lett.* **94**, 061302 (2005).
13. R. Schützhold, W. G. Unruh, *Phys. Rev. D* **66**, 044019 (2002).
14. G. E. Volovik, *J. Exp. Theor. Phys. Lett.* **76**, 240 (2002).
15. U. Leonhardt, P. Piwnicki, *Phys. Rev. Lett.* **84**, 822 (2000).
16. U. Leonhardt, *Nature* **415**, 406 (2002).
17. P. W. Milonni, *Fast Light, Slow Light and Left Handed Light* (Institute of Physics, Bristol, UK, 2004).
18. T. A. Jacobson, G. E. Volovik, *Phys. Rev. D* **58**, 064021 (1998).
19. R. Schützhold, W. G. Unruh, *Phys. Rev. Lett.* **95**, 031301 (2005).
20. G. Agrawal, *Nonlinear Fiber Optics* (Academic Press, San Diego, CA, 2001).

21. W. H. Reeves *et al.*, *Nature* **424**, 511 (2003).
22. P. Russell, *Science* **299**, 358 (2003).
23. G. 't Hooft, *Nucl. Phys. B* **256**, 727 (1985).
24. T. Jacobson, *Phys. Rev. D* **44**, 1731 (1991).
25. See the supporting material on Science Online.
26. U. Leonhardt, *Rep. Prog. Phys.* **66**, 1207 (2003).
27. *Few-Cycle Laser Pulse Generation and Its Applications*, F. X. Kärtner, Ed. (Springer, Berlin, 2004).
28. T. Brabec, F. Krausz, *Rev. Mod. Phys.* **72**, 545 (2000).
29. We are indebted to G. Agrawal, M. Dunn, T. Hänsch, A. Miller, R. Parentani, and W. Sibbett for discussions and technical support. We thank A. Podlipensky and P. Russell for measuring the dispersion of our fiber. Our work was supported by the Leverhulme Trust, Engineering and Physical Sciences Research Council, Continuous Variable Quantum Information with Atoms and Light, the Ultrafast Photonics Facility at St Andrews, and Leonhardt Group Ave.

#### Supporting Online Material

[www.sciencemag.org/cgi/content/full/319/5868/1367/DC1](http://www.sciencemag.org/cgi/content/full/319/5868/1367/DC1)  
SOM Text  
Figs. S1 to S13  
Table S1  
References and Notes

30 November 2007; accepted 24 January 2008  
10.1126/science.1153625

# Stimuli-Responsive Polymer Nanocomposites Inspired by the Sea Cucumber Dermis

Jeffrey R. Capadona,<sup>1,2,3</sup> Kadiravan Shanmuganathan,<sup>1</sup> Dustin J. Tyler,<sup>2,3</sup> Stuart J. Rowan,<sup>1,2,3,4\*</sup> Christoph Weder<sup>1,2,4\*</sup>

Sea cucumbers, like other echinoderms, have the ability to rapidly and reversibly alter the stiffness of their inner dermis. It has been proposed that the modulus of this tissue is controlled by regulating the interactions among collagen fibrils, which reinforce a low-modulus matrix. We report on a family of polymer nanocomposites, which mimic this architecture and display similar chemoresponsive mechanic adaptability. Materials based on a rubbery host polymer and rigid cellulose nanofibers exhibit a reversible reduction by a factor of 40 of the tensile modulus, for example, from 800 to 20 megapascals (MPa), upon exposure to a chemical regulator that mediates nanofiber interactions. Using a host polymer with a thermal transition in the regime of interest, we demonstrated even larger modulus changes (4200 to 1.6 MPa) upon exposure to emulated physiological conditions.

Many echinoderms share the ability to rapidly and reversibly alter the stiffness of their connective tissue (1). In the case of sea cucumbers (Fig. 1, A and B), this morphing occurs within seconds and creates considerable survival advantages (1, 2). A series of recent studies on the dermis of these invertebrates

has provided evidence that this defense mechanism is enabled by a nanocomposite structure in which rigid, high-aspect ratio collagen fibrils reinforce a viscoelastic matrix of fibrillin microfibrils (2–4). The stiffness of the tissue is regulated by controlling the stress transfer between adjacent collagen fibrils through transiently established interactions (5, 6). These interactions are modulated by soluble macromolecules that are secreted locally by neurally controlled effector cells. The dermis of the *Cucumaria frondosa* and other sea cucumber species represents a compelling model of a chemoresponsive material in which a modulus contrast by a factor of 10 (~5 to ~50 MPa) is possible (7). Intrigued by this capability and with the goal of creating new dynamic materials for biomedical applications, we set out to investigate whether nanocomposites

can be created that exhibit similar architecture and properties. The control of nanofiber interactions exploited here in solid polymer materials is similar to that observed in aqueous dispersions of poly(acrylic acid)-coated carbon nanotubes (8) or cellulose nanofibers (9), which have been shown to exhibit large viscosity changes upon variation of pH. The materials further complement other polymeric systems with morphing mechanical behavior—for example, cross-linked polymers that change cross-link density upon a change in pH or ionic concentration (10, 11).

The first series of nanocomposites studied is based on a rubbery ethylene oxide–epichlorohydrin 1:1 copolymer (EO-EPI) (Fig. 1C) into which a rigid cellulose nanofiber network was incorporated (Fig. 1, C and D). The EO-EPI matrix displays a low modulus and can accommodate the uptake of several chemical stimuli. Cellulose nanofibers, isolated for this study from the mantles of sessile sea creatures known as tunicates (12), were used as the reinforcing component. These “whiskers” exhibit high stiffness (tensile modulus ~143 GPa) (13) and dimensions at the nanometer scale (26 nm × 2.2 μm) (Fig. S1). Similar nanofibers can be obtained from a range of renewable biosources, including wood and cotton. Whiskers from tunicates were used here because their aspect ratio is higher than that of cellulose from other sources, which is advantageous for the formation of percolating architectures. Because of the high density of strongly interacting surface hydroxyl groups, cellulose whiskers have a strong tendency for aggregation (9, 14). The whisker-whisker interactions can be moderated by the introduction of sulfate surface groups (Fig. 1C), which promote dispersibility in select hydrogen-bond-forming solvents (14, 15). This balance of attractive and repulsive interactions is the key for the fabrication of cellulose-whisker nanocomposites.

<sup>1</sup>Department of Macromolecular Science and Engineering, Case Western Reserve University, Cleveland, OH 44106, USA.

<sup>2</sup>Rehabilitation Research and Development, Louis Stokes Cleveland DVA Medical Center, 10701 East Boulevard, Cleveland, OH 44106, USA.

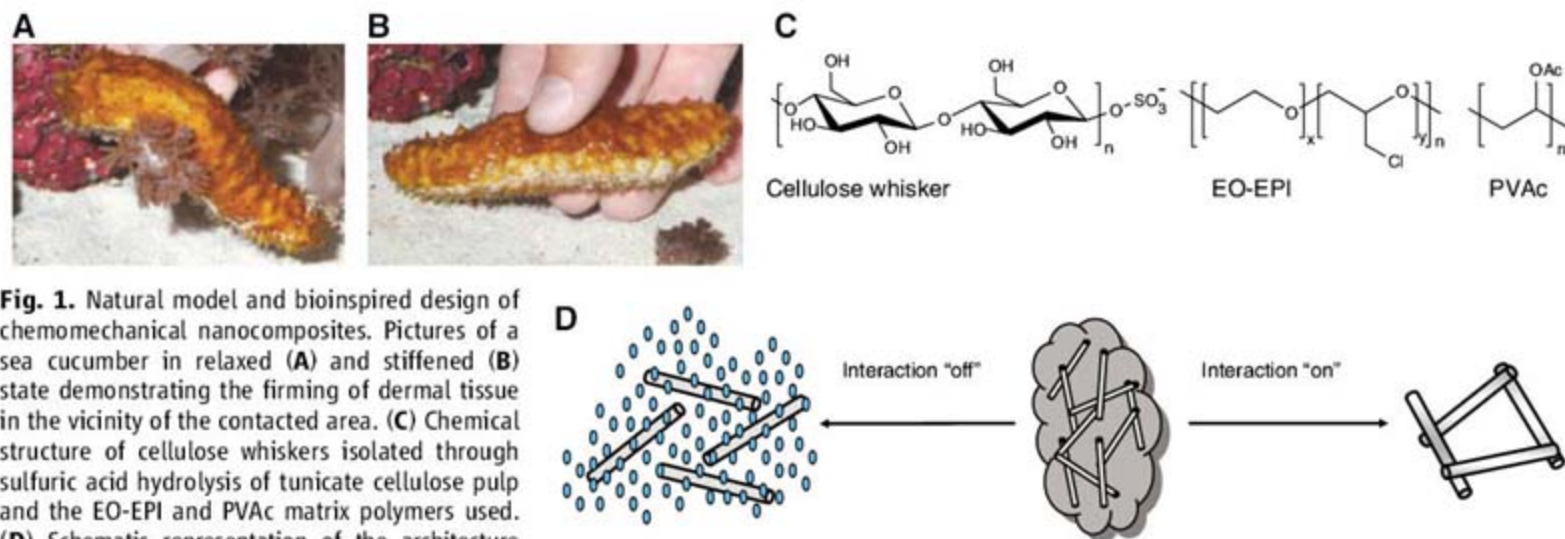
<sup>3</sup>Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH 44106, USA. <sup>4</sup>Department of Chemistry, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA.

\*To whom correspondence should be addressed. E-mail: [christoph.weder@case.edu](mailto:christoph.weder@case.edu) (C.W.); [stuart.rowan@case.edu](mailto:stuart.rowan@case.edu) (S.J.R.).

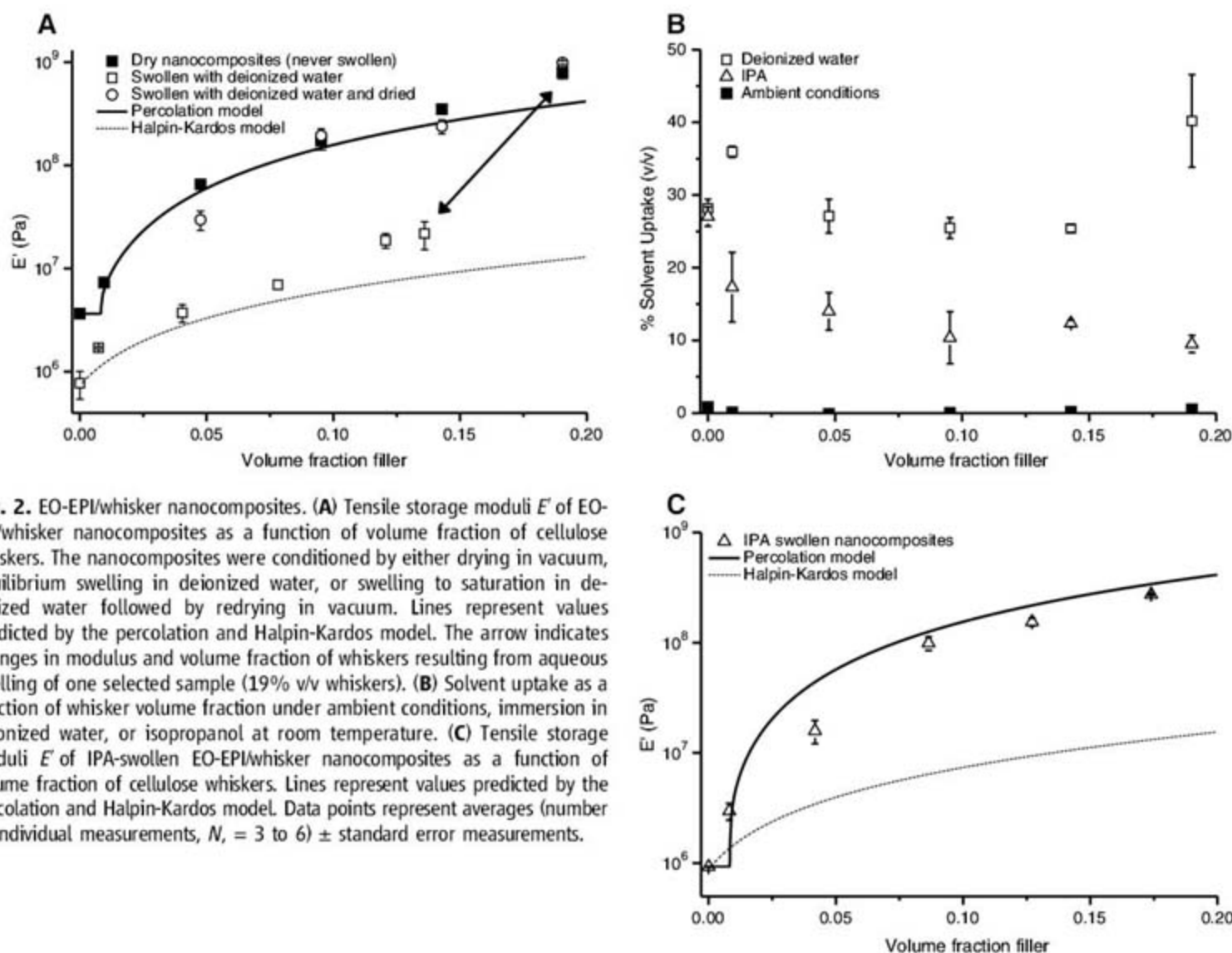
Good dispersion is achieved during processing when whisker self-interactions are “switched off” by competitive binding with a hydrogen-bond-forming solvent (Fig. 1D) (14, 15). Upon evaporation of the solvent, the interactions among the

whiskers are “switched on,” and they assemble into a percolating network. This architecture and strong interactions among the whiskers maximize stress transfer and therewith the overall modulus of the nanocomposite (13, 15, 16). Similar to the sea

cucumber dermis, it should be possible to dynamically alter the modulus of the nanocomposites through the addition or removal of a chemical regulator that, in this case, would alter the extent of hydrogen bonding of the whiskers.



**Fig. 1.** Natural model and bioinspired design of chemomechanical nanocomposites. Pictures of a sea cucumber in relaxed (A) and stiffened (B) state demonstrating the firming of dermal tissue in the vicinity of the contacted area. (C) Chemical structure of cellulose whiskers isolated through sulfuric acid hydrolysis of tunicate cellulose pulp and the EO-EPI and PVAc matrix polymers used. (D) Schematic representation of the architecture and switching mechanism in the artificial nanocomposites with dynamic mechanical properties. In the “on” state, strong hydrogen bonds between rigid, percolating nanofibers maximize stress transfer and therewith the overall modulus of the nanocomposite. The interactions are switched “off” by the introduction of a chemical regulator that allows for competitive hydrogen bonding.



**Fig. 2.** EO-EPI/whisker nanocomposites. (A) Tensile storage moduli  $E'$  of EO-EPI/whisker nanocomposites as a function of volume fraction of cellulose whiskers. The nanocomposites were conditioned by either drying in vacuum, equilibrium swelling in deionized water, or swelling to saturation in deionized water followed by redrying in vacuum. Lines represent values predicted by the percolation and Halpin-Kardos model. The arrow indicates changes in modulus and volume fraction of whiskers resulting from aqueous swelling of one selected sample (19% v/v whiskers). (B) Solvent uptake as a function of whisker volume fraction under ambient conditions, immersion in deionized water, or isopropanol at room temperature. (C) Tensile storage moduli  $E'$  of IPA-swollen EO-EPI/whisker nanocomposites as a function of volume fraction of cellulose whiskers. Lines represent values predicted by the percolation and Halpin-Kardos model. Data points represent averages (number of individual measurements,  $N = 3$  to 6)  $\pm$  standard error measurements.



EO-EPI/whisker nanocomposites were produced by solution casting from dimethylformamide (DMF), as previously reported (16), and the whisker content was varied between 0 and 19% v/v. The thermomechanical properties of these materials were established by dynamic mechanical analyses (DMA) and tensile tests. DMA temperature sweeps (figs. S2 and S3) display a glass transition temperature ( $T_g$ ) around  $-37^\circ\text{C}$  (maximum of loss tangent,  $\tan \delta$ ), which is independent of the whisker content and matches the  $T_g$  of the neat EO-EPI matrix (fig. S2). The intensity of  $\tan \delta$  decreases more than proportionally with the whisker concentration (fig. S2), which is indicative of attractive polymer-whisker interactions. Figure 2A shows the tensile storage moduli ( $E'$ ) of dry EO-EPI/whisker nanocomposites extracted from the DMA traces for a temperature of  $25^\circ\text{C}$ , that is, in the rubbery regime far above  $T_g$ .  $E'$  increased with the whisker content from  $\sim 3.7$  MPa (neat polymer) to  $\sim 800$  MPa (19% v/v whiskers). The observed reinforcement suggests the formation of a percolating nanofiber network in which stress transfer is facilitated by hydrogen-bonding between the whiskers. This hypothesis is supported by calculations obtained using a percolation model (16). Within the framework of the model, the tensile storage modulus of the nanocomposites ( $E'$ ) can be expressed as (17, 18)

$$E' = \frac{(1 - 2\psi + \psi X_r)E'_s E'_r + (1 - X_r)\psi E'_r{}^2}{(1 - X_r)E'_r + (X_r - \psi)E'_s} \quad (1)$$

with

$$\psi = X_r \left( \frac{X_r - X_c}{1 - X_c} \right)^{0.4} \quad (2)$$

where  $E'_s$  and  $E'_r$  are the experimentally determined tensile storage moduli of the neat EO-EPI (3.7 MPa) and a neat tunicate whisker film (4.0 GPa), respectively;  $\psi$  is the volume fraction of whiskers that participate in the load transfer;  $X_r$  is the volume fraction of whiskers; and  $X_c$  is the critical whisker percolation volume fraction calculated by  $0.7/A$ .  $A$  is the aspect ratio of the whiskers and has a value of 84 as determined by analysis of transmission electron microscopy (TEM) images (fig. S1). Figure 2A shows that the experimentally determined  $E'$  values of dry EO-EPI/whisker nanocomposites agree with values obtained from Eq. 1. By contrast, the data deviate strongly from the Halpin-Kardos model (fig. S4). This behavior is indicative for the formation of a percolating network of strongly interacting cellulose whiskers within the EO-EPI matrix (15, 16). This architecture is confirmed by atomic force microscopy (AFM) (Fig. 3A) and scanning electron microscopy (SEM) (Fig. 3B) images, which both show that the cellulose whiskers form a percolating network within the EO-EPI matrix. Stress-strain curves (fig. S5) reveal that the formation of a percolating network

of cellulose whiskers within the EO-EPI matrix not only affects  $E'$  but also has a considerable influence on the maximum tensile strength ( $\sigma$ ), which increased from  $0.27 \pm 0.04$  (neat EO-EPI, stress at break) to  $1.71 \pm 0.23$  MPa (14.3% v/v whiskers, stress at yield), whereas the elongation at break was reduced from  $360 \pm 20$  to  $6.7 \pm 0.8\%$  (table S1).

In view of the outstanding dispersibility of sulfate-modified cellulose whiskers in water (15), we elected to explore whether water could act as a chemical regulator for the whisker-whisker interactions in the EO-EPI/whisker nanocomposites. The atmospheric water uptake of the materials is negligible under ambient conditions, that is, if not placed in an aqueous medium (Fig. 2B). Dry EO-EPI/whisker nanocomposites were immersed in deionized water for 48 hours to achieve equilibrium swelling (Fig. 2B). Under these conditions, all compositions investigated exhibit modest aqueous swelling ( $\sim 30\%$  v/v), indicating that in the case of these compositions the water uptake is mainly governed by the matrix polymer with only minor variations due to whisker content. The tensile storage moduli for water-swollen EO-EPI/whisker nanocomposites were measured by DMA at  $25^\circ\text{C}$  in de-ionized water. A pronounced reduction of  $E'$  compared with the dry nanocomposites can be observed (Fig. 2A). The greatest mechanical contrast is seen in the case of the nanocomposite with the highest whisker content (nominally 19% v/v) (19), where  $E'$  was reduced from  $\sim 800$  to 20 MPa upon equilibrium swelling. At the same time, swelling with water leads to a significant decrease of the tensile strength ( $1.71 \pm 0.23$  to  $0.37 \pm 0.11$  MPa for a 14.3% v/v whisker nanocomposite) (fig. S5 and table S1) and an increase of the elongation at break ( $6.7 \pm 0.8$  to  $17.8 \pm 0.39\%$ ). Control experiments with the neat EO-EPI (fig. S5 and table S1) show minimal changes in tensile strength upon deionized water swelling.

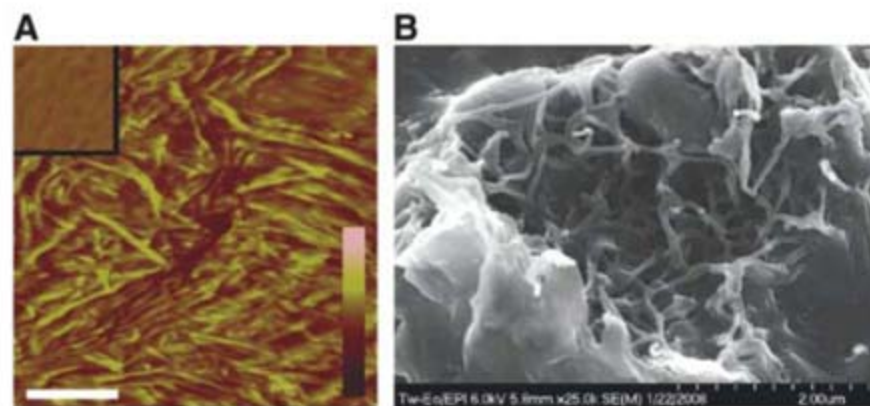
One argument that could be made against the interpretation that the observed changes in

modulus, elongation at break, and tensile strength are the result of switching off the nanofiber-nanofiber interactions is that simple swelling of the matrix alone could lead to a plasticizing effect; however, careful analysis of our data shows that this is not the case. DMA traces (fig. S2) indicate that the EO-EPI/whisker nanocomposites do not undergo any phase transition that would lead to a drop in modulus, such as cross-linked polymer hydrogels and hygroscopic polymers, which can display a decrease of the glass transition temperature upon water uptake (20). Although  $E'_s$  of the neat EO-EPI is reduced from 3.7 to 0.8 MPa upon equilibrium swelling with water (Fig. 2A), analysis in the context of the percolation model (Eqs. 1 and 2 and fig. S7) shows that a reduction of  $E'_s$  alone cannot account for a significant reduction of  $E'$ . Figure 2A also reveals that even after correcting  $X_r$  for water uptake, the percolation model no longer adequately describes  $E'$  of the water-swollen nanocomposites. By contrast, the moduli now are in much closer agreement with the Halpin-Kardos model (21), which has successfully been used to describe the modulus of nanocomposites in which the filler is homogeneously dispersed in a polymer matrix and does not display pronounced filler-filler interactions (22). The model assumes that the materials are equivalent to many layers of unidirectional plies oriented in alternating directions ( $-45^\circ$ ,  $0^\circ$ ,  $45^\circ$ , and  $90^\circ$ ), and the properties of the unidirectional reference ply are predicted by the Halpin-Tsai equations where the tensile storage modulus in the longitudinal ( $E'_L$ ) and transverse ( $E'_T$ ) directions are given by (22, 23)

$$E'_L = E'_s [1 + 2(A)\eta_L \phi_r] / (1 - \eta_L \phi_r) \quad (3)$$

$$\text{and } E'_T = E'_s [1 + 2\eta_T \phi_r] / (1 - \eta_T \phi_r) \quad (4)$$

Thus, all data indicate that the stiffness reduction achieved in the EO-EPI/whisker nanocomposites is related to the decoupling of the stress-transferring rigid nanofiber network upon



**Fig. 3.** Morphology of EO-EPI/whisker nanocomposites. **(A)** Representative AFM phase image of an ultramicrotomed nanocomposite containing 9.5% v/v whiskers in EO-EPI. The inset shows an AFM phase image of a neat EO-EPI reference (horizontal scale bar = 500 nm, vertical scale bar = phase shift 120 to  $0^\circ$ ). The samples were briefly (10 s) immersed in tetrahydrofuran and rinsed with IPA to partially dissolve the polymer at the surface of the sample and to expose the inner structure of the films. **(B)** Representative SEM image of an untreated nanocomposite containing 9.5% v/v whiskers in EO-EPI.

introduction of water as a competitive hydrogen-bonding agent. Consistent with the proposed mechanism, the switching is fully reversible: The materials adapted their original stiffness upon drying (Fig. 2A).

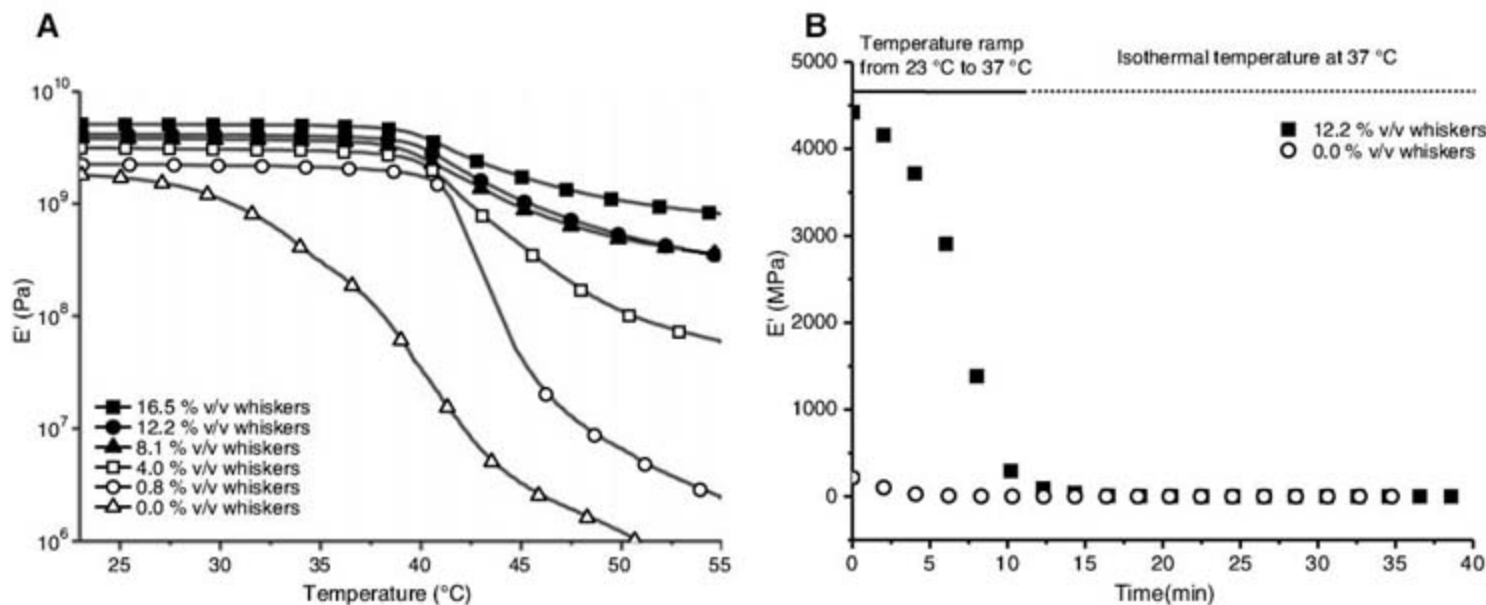
To investigate the specificity of the switching mechanism, we investigated the effect of isopropanol (IPA) as the swelling agent. IPA was selected because it swells neat EO-EPI to a similar degree as water (Fig. 2B) but is incapable of dispersing cellulose whiskers (16). The nanocomposites swelled upon immersion in IPA (Fig. 2B) to a level similar to that of the composites in water; however  $E'$  barely changed in comparison to the dry state (Fig. 2C), and the data fit the percolation model. This result confirms that the chemomechanical response is largely a result of disruption of the whisker-whisker interactions and not just plasticization of the material. By contrast, EO-EPI is plasticized considerably upon IPA swelling ( $E'$  drops from 3.6 to 0.93 MPa). This contrast highlights the most important advantage of the nanocomposite approach over simple plasticization of a neat polymer. Although plasticization through solvent uptake, which is inherent to the latter, is a nonspecific process, the responsive nanocomposites can be designed to display a response that is specific to the nature of the stimulus. In addition, the nanocomposite approach provides the ability to increase the initial stiffness and strength of the material and allows for the use of host polymers that have no thermal transition in the temperature regime of interest, such as the EO-EPI matrix used here.

We are interested in exploiting dynamic mechanical materials in biomedical applications, specifically as adaptive substrates for intracortical microelectrodes. These implants have the ability to record brain unit activity (24). Brain-machine interfaces that rely on these electrodes provide

solutions to medical conditions such as Parkinson's disease, stroke, and spinal chord injuries (25). One problem with current microelectrodes is that the signal quality usually degrades within a few months, making chronic applications challenging (26). One hypothesis for the cause of possible failure, especially in recording applications, is that the micromotion of rigid electrodes within the soft cortical tissue chronically inflicts trauma on the surrounding neurons (27). We hypothesize that a mechanically adaptive electrode could alleviate this problem, and we are thus interested in designing devices that are initially rigid to allow for penetration of the pia mater during implantation (28) but that soften slowly and without excessive expansion upon implantation in response to the chemical environment within the brain (for an emulation, see Movie S1). For this application, an initial  $E'$  of  $>4$  GPa is desirable to allow for the insertion of an electrode with typical dimensions into the cortex (29). Because EO-EPI/whisker nanocomposites display a substantially lower  $E'$ , we sought to combine the switching mechanism with a chemically influenced thermal transition. We discovered that nanocomposites based on poly(vinyl acetate) (PVAc) (Fig. 1C) and cellulose whiskers display such a "dual" responsive behavior. Our data show that, upon exposure to physiological conditions, the materials undergo a phase transition; in addition, the reinforcing whisker network is disassembled. DMA experiments (Fig. 4A and fig. S8) reveal that the neat PVAc displays a  $T_g$  around 42°C, that, just above physiological temperature,  $E'_s$  of the neat polymer is considerably reduced upon heating from room temperature (1.8 GPa at 23°C) to above  $T_g$  (0.39 MPa at 56°C) (this corresponds to  $T_g + 16^\circ\text{C}$  and marks the temperature at which  $E'_s$  is starting to level off). As evidenced by DMA data,

the introduction of cellulose whiskers into PVAc has only a minimal influence on  $T_g$  in the dry state (Fig. 4A and fig. S8). The thermal transition is sharpened, and the temperature at which  $E'$  begins to drop is increased from  $\sim 25$  to  $>40^\circ\text{C}$ . For certain biomedical applications, this effect is very desirable, because it prevents the thermally induced softening of the material just upon exposure to body temperature. As a consequence of the already rather high stiffness of the glassy PVAc matrix, only a modest reinforcement is observed for the nanocomposites below  $T_g$  ( $E' = 5.1$  GPa with 16.5% v/v whiskers) (fig. S8). However, a dramatic effect is observed above  $T_g$ , where  $E'$  is increased from 1.0 MPa for the neat polymer matrix up to 814 MPa with 16.5% v/v whiskers (at 56°C). The experimental data above  $T_g$  match well with the percolation model (fig. S8), which indicates that also in this series a percolating network of strongly interacting whiskers is formed (15, 16, 18). The nanocomposites demonstrate significant swelling in both deionized water and artificial cerebrospinal fluid (ACSF). The solvent uptake increases with increasing whisker content and temperature (fig. S9), lowers the  $T_g$  to below physiological temperature (19 to 23°C) (fig. S10), and reduces  $E'$  dramatically. For example, the  $E'$  of a 16.5% v/v whisker nanocomposite above  $T_g$  is reduced from 814 MPa (dry) to 10.8 MPa (water swollen; data are for 56 and 37°C, respectively, that is, 16°C above the respective  $T_g$ ). As for the water-swollen EO-EPI/whisker nanocomposites, the moduli of the ACSF swollen PVAc/whisker nanocomposites are better described by the Halpin-Kardos than the percolation model (fig. S8), again indicative of decoupling of the stress-transferring nanofiber network upon introduction of water.

Exposure to brain tissue, simulated here by immersing the samples into ACSF and heating to



**Fig. 4.** PVAc/whisker nanocomposites. (A) Tensile storage moduli  $E'$  of dry films of PVAc/whisker nanocomposites as a function of temperature. The nanocomposites contain between 0 and 16.5% v/v whiskers. (B) Time-dependent modulus decrease of neat PVAc and a 12.2% v/v PVAc/

whisker nanocomposite upon immersion into ACSF and increasing the temperature from 23°C to 37°C. Lines represent time required for temperature to increase from 23°C to 37°C and isothermal control at 37°C.

a physiological temperature of 37°C at -2°C/min (30) (Movie S1), leads to a pronounced reduction of  $E'$ . Whereas the neat PVAc (dry  $E'_s = 1.8$  GPa at 25°C) instantly softens under these conditions (Fig. 4B), the  $E'$  of the whisker-reinforced nanocomposites (see Fig. 4B for a 12.2% v/v nanocomposite) is reduced slowly over a period of 15 min. The whisker-reinforced nanocomposite displays a much higher dry  $E'$  (4.2 GPa at 25°C) than the neat PVAc, but both materials reach nearly identical moduli upon immersion in ACSF at 37°C (1.6 MPa).

Our data support a simple and versatile strategy for the creation of polymer nanocomposites, whose architecture and mechanical adaptability closely mimic the proposed architecture and observed response of the sea cucumber demis. The mechanical properties of these chemoresponsive materials can selectively and reversibly be controlled through the formation and decoupling of a three-dimensional network of well-individualized nanofibers in response to specific chemical triggers. It will be interesting to explore whether the framework can be adapted to nonchemical triggers, for example, optical or electrical stimuli.

#### References and Notes

1. T. Heinzelner, J. Nebelsick, Eds., *Echinoderms* (Taylor & Francis, London, 2004).
2. T. Motokawa, *Comp. Biochem. Physiol. B* **109**, 613 (1994).
3. F. A. Thurmond, J. A. Trotter, *J. Exp. Biol.* **199**, 1817 (1996).
4. I. C. Wilkie, *J. Exp. Biol.* **205**, 159 (2002).
5. J. A. Trotter, T. J. Koob, *Matrix Biol.* **18**, 569 (1999).
6. G. K. Szulgít, R. E. Shadwick, *J. Exp. Biol.* **203**, 1539 (2000).
7. J. A. Trotter et al., *Biochem. Soc. Trans.* **28**, 357 (2000).
8. J. C. Grunlan, L. Liu, Y. S. Kim, *Nano Lett.* **6**, 911 (2006).
9. M. M. de Souza Lima, R. Borsali, *Macromol. Rapid Commun.* **25**, 771 (2004).
10. J. A. Jaber, J. B. Schlenoff, *J. Am. Chem. Soc.* **128**, 2940 (2006).
11. D. M. Loveless, S. L. Jeon, S. L. Craig, *J. Mater. Chem.* **17**, 56 (2007).
12. P. S. Belton, S. F. Tanner, N. Cartier, H. Chanzy, *Macromolecules* **22**, 1615 (1989).
13. A. Sturcova, J. R. Davies, S. J. Eichhorn, *Biomacromolecules* **6**, 1055 (2005).
14. O. van den Berg, J. R. Capadona, C. Weder, *Biomacromolecules* **8**, 1353 (2007).
15. M. A. S. Azizi Samir, F. Alloin, A. Dufresne, *Biomacromolecules* **6**, 612 (2005).
16. J. R. Capadona et al., *Nat. Nanotech.* **2**, 765 (2007).
17. M. Takayanagi, S. Uemura, S. Minami, *J. Polym. Sci. C* **5**, 113 (1964).
18. N. Ouali, J. Y. Cavallé, J. Pérez, *J. Plast. Rubber Comp. Process. Appl.* **16**, 55 (1991).
19. Swelling increased the volume of the nanocomposites and reduced the volume fraction of whiskers,  $X_i$ . For example, when a nanocomposite with  $X_i = 19\%$  w/v was swollen with water,  $X_i$  decreased to 14%. The representation of data in Fig. 2A considers this effect to allow for analysis by the Halpin-Kardos model. A direct comparison of dry versus wet composites for the same fiber loading is shown in fig. S6.
20. J. Kunzelman, B. R. Crenshaw, C. Weder, *J. Mater. Chem.* **17**, 2989 (2007).
21. J. C. Halpin, J. L. Kardos, *J. Appl. Phys.* **43**, 2235 (1972).
22. P. Hajji, J. Y. Cavallé, V. Favier, C. Gauthier, G. Vigier, *Polym. Compos.* **17**, 612 (1996).
23.  $\nu_s = [(E_{ii} / E'_s) - 1] / [(E_{ii} / E'_s) + 2A]$ , and  $\eta_t = [(E_{ii} / E'_s) - 1] / [(E_{ii} / E'_s) + 2]$ .  $A$  is the aspect ratio of the whiskers,  $\phi$  is equal to the volume fraction of the phase, and the subscripts  $s$  and  $r$  represent the soft phase and the rigid phase, respectively.  $E_{ii}$  is the longitudinal Young's Modulus (130 GPa), and  $E_{ii}$  is the transverse Young's Modulus (5 GPa) of an individual cellulose whisker (22). To determine the tensile storage modulus of the isotropic nanocomposite ( $E'$ ),  $E'_s$  and  $E'_r$  must be

inserted into one equation using the Halpin-Kardos model:  $E' = 4U_5(U_1 - U_2)/U_1$  where  $U_1 = 1/8(3Q_{11} + 3Q_{22} + 4Q_{66})$ ;  $U_2 = 1/8(Q_{11} + Q_{22} - 2Q_{12} + 4Q_{66})$ ;  $Q_{11} = E'/(1 - \nu_{12}\nu_{21})$ ;  $Q_{22} = E'_r(1 - \nu_{12}\nu_{21})$ ;  $Q_{12} = \nu_{12}Q_{22} = \nu_{21}Q_{11}$ ;  $Q_{66} = G_{12}$ ;  $\nu_{12} = \phi_s\nu_s + \phi_r\nu_r = 0.3$ ;  $G_{12} = G_s(1 + \eta\phi_s)/(1 - \eta\phi_s)$ ;  $\eta = (G_r/G_s - 1)/(G_r/G_s + 1)$ ,  $\nu$  is the Poisson's ratio,  $G$  is the shear modulus, and  $G_r = 1.77$  GPa.

24. D. M. Taylor, S. I. H. Tillery, A. B. Schwartz, *Science* **296**, 1829 (2002).
25. A. B. Schwartz, *Annu. Rev. Neurosci.* **27**, 487 (2004).
26. R. Biran, D. C. Martin, P. A. Tresco, *J. Biomed. Mater. Res.* **82A**, 169 (2007).
27. W. L. C. Rutten, *Annu. Rev. Biomed. Eng.* **4**, 407 (2002).
28. D. H. Szarowski et al., *Brain Res.* **983**, 23 (2003).
29. K. Najafi, J. F. Hetke, *IEEE Trans. Biomed. Eng.* **37**, 474 (1990).
30. Materials and methods are available as supporting material on Science Online.
31. We thank F. Carpenter for the photography of the sea cucumber and L. McCorkle, J. Johnson, and M. Hitomi for assistance with the SEM, AFM, and TEM experiments, respectively. Financial support from DuPont (Young Professor Award to C.W.), the L. Stokes Cleveland VAMC Advanced Platform Technology Center, an Ohio Innovation Incentive Fellowship (to K.S.), the Department of Veterans Affairs Associate Investigator Career Development Program (to J.C.), and the National Institutes of Health are gratefully acknowledged. The authors declare that they have no competing financial interest.

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5868/1370/DC1  
Materials and Methods  
Figs. S1 to 10  
Table S1  
References  
Movie S1

6 November 2007; accepted 1 February 2008  
10.1126/science.1153307

## Heterogeneous Nucleation Experiments Bridging the Scale from Molecular Ion Clusters to Nanoparticles

Paul M. Winkler,<sup>1</sup> Gerhard Steiner,<sup>1</sup> Aron Vrtala,<sup>1</sup> Hanna Vehkamäki,<sup>2</sup> Madis Noppel,<sup>3</sup> Kari E. J. Lehtinen,<sup>4</sup> Georg P. Reischl,<sup>1</sup> Paul E. Wagner,<sup>1</sup> Markku Kulmala<sup>2\*</sup>

Generation, investigation, and manipulation of nanostructured materials are of fundamental and practical importance for several disciplines, including materials science and medicine. Recently, atmospheric new particle formation in the nanometer-size range has been found to be a global phenomenon. Still, its detailed mechanisms are mostly unknown, largely depending on the incapability to generate and measure nanoparticles in a controlled way. In our experiments, an organic vapor (*n*-propanol) condenses on molecular ions, as well as on charged and uncharged inorganic nanoparticles, via initial activation by heterogeneous nucleation. We found a smooth transition in activation behavior as a function of size and activation to occur well before the onset of homogeneous nucleation. Furthermore, nucleation enhancement for charged particles and a substantial negative sign preference were quantitatively detected.

Condensational growth, evaporation, and heterogeneous chemistry are important phenomena in materials science, fluid dynamics, aerosol physics and technology, and atmospheric chemistry, including cloud microphysics and cloud chemistry. A prerequisite for the start of condensation is homogeneous nu-

cleation of new particles or the activation of preexisting particles by heterogeneous nucleation. The latter can occur either on ions, soluble particles, or insoluble particles, and is energetically easier than homogeneous nucleation (1). Both particle formation processes are of fundamental as well as practical importance and

have been the subject of investigations for more than a century (2). Important examples representing the different processes are given by the use of the Wilson cloud chamber (3) in high-energy physics for the case of ions, the formation of cloud droplets in the troposphere for the case of soluble particles, and the occurrence of ice nucleation for the case of insoluble seed particles (4). Atmospheric observations suggest that the initial formation and growth are two uncoupled processes (5–7), and therefore the activation mechanism of small clusters is of vital importance. Understanding the formation and initial growth processes in detail is also crucial to control the production of nanomaterials (8).

In this paper, we present experimental results for the activation of molecular cluster ions, charged and neutral clusters, and nanometer-size particles having almost monodisperse size distributions,

<sup>1</sup>Fakultät für Physik, Universität Wien, Boltzmannngasse 5, A-1090 Wien, Austria. <sup>2</sup>University of Helsinki, Department of Physical Sciences, Post Office Box 64, 00014 University of Helsinki, Finland. <sup>3</sup>Institute of Physics, University of Tartu, 18 Ülikooli Street, 50090 Tartu, Estonia. <sup>4</sup>Department of Physics, University of Kuopio and Finnish Meteorological Institute, Post Office Box 1627, 70211 Kuopio, Finland.

\*To whom correspondence should be addressed. E-mail: markku.kulmala@helsinki.fi

with mean diameters ranging from 1 to 24 nm and geometric SDs between 1.021 and 1.058. Particle number concentrations were between 2000 and 10,000  $\text{cm}^{-3}$ . We used insoluble inorganic seed aerosol ( $\text{WO}_x$ ) and condensable organic vapor (*n*-propanol) to determine the dependence of activation diameter on charging state and vapor saturation ratio. For comparison, other insoluble inorganic seed aerosols [ $\text{Ag}$  and  $(\text{NH}_4)_2\text{SO}_4$ ] were also used (9). The range of particle sizes studied—together with the fact that the particles are, in practice, monodispersed—is the main novelty of our work: To our knowledge, this is the first time that activation of neutral particles has been studied for particles smaller than 3 nm in diameter.

The experimental system (9) includes a source of monodispersed particles and a vapor generation unit. Vapor supersaturation is achieved by adiabatic expansion in a computer-controlled thermostated expansion chamber [the size-analyzing nuclei counter (SANC)]. Droplet growth is observed by means of the constant-angle Mie scattering (CAMS) detection method (10). Details of the experimental system are presented elsewhere (11).

We generated well-defined nearly monodispersed nanoparticles using different types of particle generators in combination with an electrostatic classification system capable of selecting particles with specified charging state and mobility diameters down to and even below 1 nm. Ions were obtained from a radioactive source in a  $^{241}\text{Am}$  charger. Charged particles leaving the classifier were either passed through a  $^{241}\text{Am}$  neutralizer to obtain neutral particles or passed by the neutralizer to preserve charging properties. In order to remove ions, which were produced in the neutralizer, from the aerosol flow, we applied an ion trap. *n*-Propanol vapor was added to the system by controlled injection of a liquid beam and subsequent quantitative evaporation in a heated unit. Those particles inducing heterogeneous nu-

cleation in the SANC expansion chamber at the vapor supersaturations considered lead to the formation and growth of liquid droplets, which were optically detected.

We measured heterogeneous nucleation (activation) probabilities using the above described SANC-CAMS method. For all experiments reported in this paper, the nucleation temperature was kept constant at about 275 K. For each *n*-propanol vapor saturation ratio, the fraction of activated particles relative to total particle number concentration was determined, resulting in a nucleation-activation probability curve. Experimental nucleation probabilities, determined for molecular ions and nanometer-size particles, are shown in Fig. 1. The diameters of the seed objects range from 4 nm down to 0.9 nm. The smaller the size, the higher the saturation ratio that is needed for activation, reflecting the well-known curvature effect first observed by Thomson (12). As can be seen, at 4 nm, the nucleation probability curves are mainly overlapping, whereas with decreasing particle size, the curves split up, indicating that charge effects increasingly promote the nucleation process. Negatively charged particles are found to require smaller saturation ratios to activate for nucleation than positively charged particles and neutral ones.

Each probability curve can be represented by the corresponding onset saturation ratio (i.e., the saturation ratio where 50% of the particles of a certain size are activated). Experimental onset saturation ratios as functions of mobility diameter, together with the Kelvin diameter and the homogeneous nucleation limit, are shown in Fig. 2. The Kelvin diameter corresponds to the size at which the vapor and aerosol particles, whose surfaces are coated with *n*-propanol, are in equilibrium (9). In principle, if the particle size is larger than the Kelvin diameter, the aerosol particles will grow; if the particle size is smaller, they will evaporate. However, in our experiments, all particles considered are activated and start to grow at

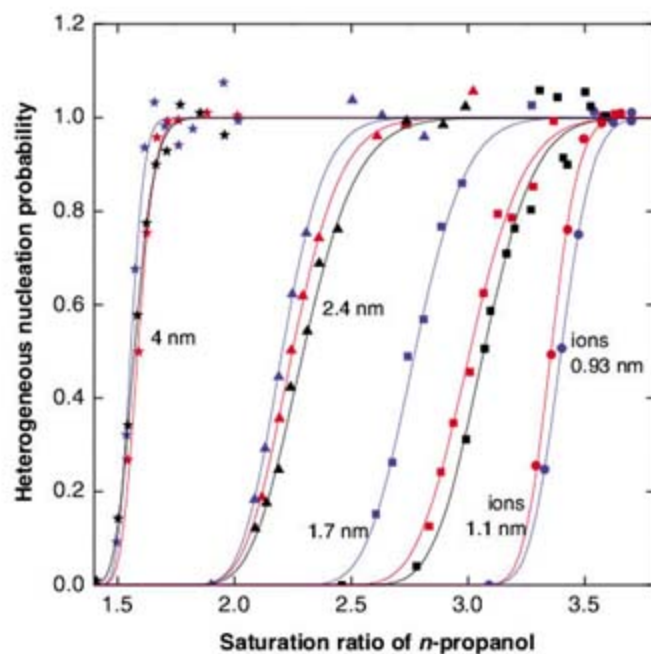
sizes clearly smaller than those indicated by the Kelvin equation and at saturation ratios well below the onset of homogeneous nucleation.

For particles with diameters around 4 nm and above, nucleation behavior was found to be independent of charging state. However, as seen already in Fig. 1, charge effects become increasingly important with decreasing particle size. Differences between heterogeneous nucleation on charged and neutral particles were experimentally observed, starting at diameters below 3 nm. Clearly, a “sign effect” can be seen, meaning that negatively charged particles appear to be more easily activated for growth than positively charged particles. Smaller saturation ratios are generally required to activate charged particles for nucleation. For example, at a saturation ratio of 2.75, the activation size is 1.4 nm for negative clusters and 1.8 nm for neutral ones. A smooth transition from molecular ions to particles is observed, thereby bridging the scale from molecular ion clusters to nanoparticles. Consistently, for ions, a similar sign effect has been found as for charged particles. Recently, Nadykto *et al.* (13) have proposed that the sign can be reproduced by quantum chemical methods and may be positive or negative, depending on the nature of the ions considered.

In Fig. 2, we also compare our experimental findings with an onset activation curve predicted by heterogeneous nucleation theory (9, 14). In heterogeneous nucleation, the critical cluster is formed on a preexisting surface, and only part of the vapor-liquid surface needs to be built from scratch, because part of it can be borrowed from the preexisting interface; this fact decreases the energy barrier of cluster formation as compared with homogeneous nucleation (1, 11). The angle between the preexisting surface and the surface of the nucleating cluster is called the contact angle  $\theta$ . The value that we used for the contact angle was  $0^\circ$ , corresponding to the recent observations for *n*-propanol droplets nucleating on an Ag surface (11). Substantially larger contact angles would lead to a predicted activation diameter being larger than the Kelvin diameter (1), which is in contrast to our experimental findings; thus, we conclude that the particles are, in practice, totally wettable to the *n*-propanol vapor. In atmospheric and materials science, heterogeneous nucleation is typically ignored. This is mainly because no proper experimental data have so far existed at particle sizes small enough to be relevant to formation and initial growth. Our results, however, show that the theory developed by Fletcher 50 years ago predicts the observed onset activations for neutral particles and clusters exceedingly well, even at the size range of small molecular clusters.

The effect of charge on the onset activation curves can be estimated by ion-induced nucleation theory—with Gibbs free energies calculated based on the Thomson theory (15)—together with the Fletcher theory (9), hereafter referred to as the “combined” theory. The onset activation satura-

**Fig. 1.** Heterogeneous nucleation probabilities versus vapor saturation ratio for nucleation of *n*-propanol on ions as well as charged and uncharged  $\text{WO}_x$  clusters of different diameters:  $\text{WO}_x$ , 4 nm (stars);  $\text{WO}_x$ , 2.4 nm (triangles);  $\text{WO}_x$ , 1.7 nm (squares); positive ions, 1.1 nm (red circles); and negative ions, 0.93 nm (blue circles). Colors refer to different charging states: blue (negatively charged), red (positively charged), and black (neutral). Solid lines are shown to guide the eye.



tion ratio as a function of diameter, estimated by the combined theory, is also superimposed in Fig. 2, showing an even lower supersaturation required for activation. The combined theory provides no explanation why negative particles activate before positive ones of the same size. However, the difference between the Fletcher theory (heterogeneous nucleation) and the combined theory (heterogeneous and ion-induced nucleation) seems to provide a useful estimate for the difference in onset activation supersaturations between charged and neutral particles.

The use of the recently developed heterogeneous nucleation theorem (16) will provide an estimation of the number of molecules in the nucleating cluster. In the experiments presented here, the number is around 20 to 25 molecules. The ratio of the experimental activation diameter to the Kelvin diameter is almost size-independent and is around 0.65. Our experimental technique can thus be used to produce particles, with a well-defined surface layer consisting of a small number of molecules, for future technical and industrial purposes.

So far, it has been generally accepted (1, 17) that activation of (neutral) particles is predicted by the Kelvin equation. The fact that particles are activated at much smaller sizes than indicated by the Kelvin equation is thus unexpected; apparently, heterogeneous nucleation has been typically forgotten or at least ignored, and the Fletcher theory has not been applied to small clusters or nanoparticles. The latest findings will give insight to some aspects of materials science, atmospheric particle formation, and especially to measurement techniques of small aerosol particles. Based on laboratory experiments, atmospheric observations, and modeling (18–20), it is often predicted that homogeneous nucleation will produce particles with diameters around 1 to 2 nm. Indeed, Strey *et al.* (21) have shown that, for the case of homogeneous nucleation (i.e., in the absence of seed particles),

the critical cluster size agrees well with the Kelvin equation; this indicates that the Kelvin equation remains generally valid for critical clusters down to the size range of 1 to 2 nm. Our results show that, for example, if the Kelvin diameter is 2 nm, then the activation diameter is around 1.2 nm; this finding indicates that all seed particles with diameters larger than 1.2 nm are capable of acting as nuclei for condensation, which is also consistent with the recent observation of size-dependent aerosol growth (6). As described by the Fletcher theory (9, 14), small embryos have a thermodynamic barrier to cross, but critical size clusters are nevertheless formed on seed particles as a result of statistical fluctuations. They reach the critical size and become stable for growth when their radius coincides with the Kelvin curvature. The work of formation of an embryo on a seed particle is always smaller than the work of formation of a Kelvin-size embryo in homogeneous nucleation, even if the seed particle diameter is below the Kelvin size. This is manifested by our finding that organic vapor will condense much easier on even the smallest preexisting seed particles than form new particles.

In nanomaterial production, there are several examples of experiments where the detailed particle formation, activation, and/or initial growth mechanisms are unknown, and modeling attempts have been unsuccessful (22, 23). These modeling efforts are typically based on the application of classical homogeneous nucleation theory with some adjustable factor, together with coagulation and condensation models for the growth. The heterogeneous nucleation-activation stage is typically neglected, and condensation is assumed to start at the Kelvin limit. The present study proposes one severe deficiency in these approaches, because we clearly observe that activation and condensational growth can start at sizes significantly smaller than previously expected from the Kelvin equation.

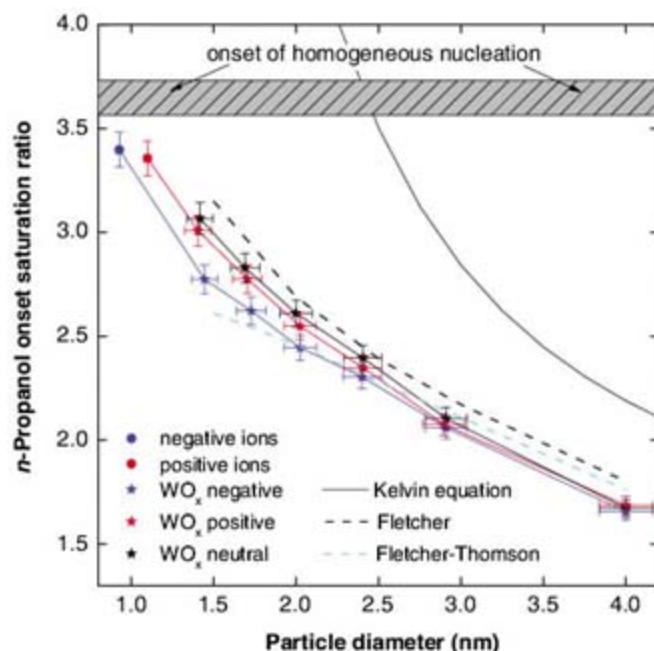
One of the main implications of this study is that the detection limit of condensation particle counters (CPCs) can be extended to considerably smaller sizes (24, 25). The operating principle of these devices is based on activation of particles in supersaturated vapors and consequent growth to sizes detectable by optical methods. A major concern has been possible homogeneous nucleation of new particles when the application of high supersaturations is required to activate nanoparticles. However, this study shows that, rather than nucleating homogeneously to form new particles, the vapors tend to nucleate heterogeneously and further condense onto the preexisting particles at sizes much smaller than believed in the past. This fact has already been used in atmospheric studies (1).

In the literature, there are some laboratory studies concerning the effect of ions on the nucleation rate. However, they are not typically related to activation of existing particles or clusters. For example, Kim *et al.* (26) investigated homogeneous and ion-induced nucleation in the ternary  $\text{NH}_3/\text{SO}_2/\text{H}_2\text{O}/\text{air}$  mixture. As a result of their study, they proposed that the main particle production mechanism is homogeneous nucleation of  $(\text{NH}_4)_2\text{SO}_4$  molecules produced by the  $\text{H}_2\text{SO}_4\text{-NH}_3$  reaction. Several years ago, Gamero-Castaño and de la Mora (17) proposed clusters as “impurities in the gas phase” using their CPC. Their study focused on the activation of ions and charged nanoclusters with the use of a CPC, but did not include any comparison between charged and neutral particles.

The presence of ions, and electric charge on particles, will enhance not only the activation probabilities but also the growth rates of very small (nanometer-size) aerosol particles and air ions (27, 28). The condensing vapors may exhibit dipole nature and can thus be electrostatically attracted to charged particles. This effect decreases rapidly with increasing particle size, which means that if this mechanism dominates the growth of freshly nucleated particles, the particle growth rate should decrease as a function of the particle size. However, no sign of such a growth dependence has been reported. Although the condensation enhancement factor caused by the presence of electric charges varies between the different theories, all of the theories predict a fairly similar size dependence for this effect (6). However, according to our present study, the possible difference in growth has a much smaller effect on atmospheric aerosol formation than the activation probability. Therefore, we can conclude that the key process in atmospheric investigations is the activation of preexisting clusters and nanoparticles.

From the point of view of atmospheric aerosols and their climate interaction, our observation allows a more accurate description of cluster activation leading to aerosol formation. In atmospheric processes, several different organic compounds are undoubtedly participating in the activation process.

**Fig. 2.** Experimental onset saturation ratio as a function of seed particle mobility diameter for ions as well as for charged and uncharged  $\text{WO}_x$  particles. The Kelvin diameter (solid black line) and the homogeneous nucleation onset [cross-hatched shaded area (29)] are shown for comparison. We also show the predictions by the Fletcher theory for heterogeneous nucleation on uncharged particles (dashed black line) and the combined theory accounting for the charge of the seed particles (dashed cyan line) (9). Error bars indicate the measurement uncertainty when measuring saturation ratio and particle diameter.



## References and Notes

- M. Kulmala et al., *J. Aerosol Sci.* **38**, 289 (2007).
- C. T. R. Wilson, *Philos. Trans. R. Soc. London Ser. A* **189**, 265 (1897).
- N. N. Das Gupta, S. K. Ghosh, *Rev. Mod. Phys.* **18**, 225 (1946).
- H. R. Pruppacher, J. D. Klett, *Microphysics of Clouds and Precipitation* (Kluwer Academic, Dordrecht, Netherlands, 1997).
- M. Kulmala, L. Pirjola, J. M. Mäkelä, *Nature* **404**, 66 (2000).
- M. Kulmala et al., *Atmos. Chem. Phys.* **4**, 2553 (2004).
- M. Kulmala, K. E. J. Lehtinen, A. Laaksonen, *Atmos. Chem. Phys.* **6**, 787 (2006).
- T. T. Kodas, M. J. Hampden-Smith, *Aerosol Processing of Materials* (Wiley-VCH, New York, 1999).
- Detailed information on experimental methods, further results, and theory is available as supporting material on Science Online.
- P. E. Wagner, *J. Colloid Interface Sci.* **105**, 456 (1985).
- P. E. Wagner, D. Kallier, A. Virtala, A. Lauri, M. Kulmala, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **67**, 021605 (2003).
- W. Thomson, *Proc. R. Soc. Edinburgh* **7**, 63 (1870).
- A. B. Nadykto, A. A. Natsheh, F. Yu, K. V. Mikkelsen, J. Ruuskanen, *Phys. Rev. Lett.* **96**, 125701 (2006).
- N. Fletcher, *J. Chem. Phys.* **29**, 572 (1958).
- M. Noppel, H. Vehkamäki, M. Kulmala, *J. Chem. Phys.* **119**, 10733 (2003).
- H. Vehkamäki et al., *J. Chem. Phys.* **126**, 174707 (2007).
- M. Gamero-Castaño, J. F. de la Mora, *J. Aerosol Sci.* **31**, 757 (2000).
- M. Kulmala, *Science* **302**, 1000 (2003).
- M. Kulmala et al., *J. Aerosol Sci.* **35**, 143 (2004).
- D. Kashchiv, *Nucleation: Basic Theory with Applications* (Butterworth-Heinemann, Oxford, 2000).
- R. Strey, P. E. Wagner, Y. Viisanen, *J. Phys. Chem.* **98**, 7748 (1994).
- M. Wilck, F. Stratmann, *J. Aerosol Sci.* **28**, 959 (1997).
- U. Backman, J. K. Jokiniemi, A. Auvinen, K. E. J. Lehtinen, *J. Nanoparticle Res.* **4**, 325 (2002).
- P. H. McMurry, *Atmos. Environ.* **34**, 1959 (2000).
- M. Kulmala, K. E. J. Lehtinen, L. Laakso, G. Mordas, K. Hämeri, *Boreal Environ. Res.* **10**, 79 (2005).
- T. O. Kim, T. Ishida, M. Adachi, K. Okuyama, J. H. Seinfeld, *Aerosol Sci. Technol.* **29**, 111 (1998).
- F. Yu, R. P. Turco, *Geophys. Res. Lett.* **27**, 883 (2000).
- A. A. Lushnikov, M. Kulmala, *Eur. Phys. J. D* **29**, 345 (2004).
- J. Wedekind, K. Iland, P. E. Wagner, R. Strey, in *Nucleation and Atmospheric Aerosols 2004*, M. Kasahara, M. Kulmala, Eds. (Kyoto Univ. Press, Kyoto, 2004), pp. 49–52.
- This work was supported by the Austrian Science Foundation (Project No. P16958-N02 and P19546-N20), the Estonian Science Foundation (Grants 6223 and 6988), and the Academy of Finland. The authors declare no competing interests. Authors' contributions statement: experiments (P.M.W. and P.E.W.), nanoparticle generation (G.S., G.P.R., and P.M.W.), data analysis (P.M.W. and A.V.), theory and model calculations (M.K., M.N., H.V., and K.E.J.L.), writing (M.K., P.E.W., P.M.W., G.S., K.E.J.L., M.N., and H.V.).

## Supporting Online Material

[www.sciencemag.org/cgi/content/full/319/5868/1374/DC1](http://www.sciencemag.org/cgi/content/full/319/5868/1374/DC1)

SOM Text

Figs. S1 to S5

Table S1

References

9 August 2007; accepted 24 January 2008

10.1126/science.1149034

# Age and Evolution of the Grand Canyon Revealed by U-Pb Dating of Water Table–Type Speleothems

Victor Polyak,\* Carol Hill, Yemane Asmerom

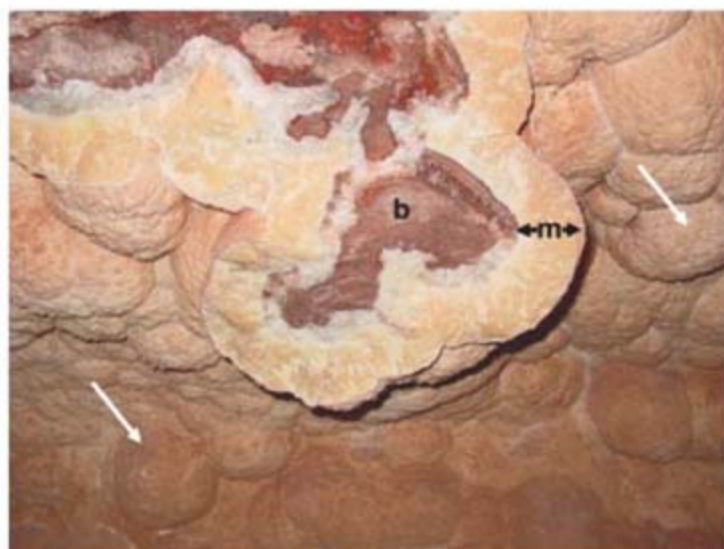
The age and evolution of the Grand Canyon have been subjects of great interest and debate since its discovery. We found that cave mammillaries (water table indicator speleothems) from nine sites in the Grand Canyon showed uranium-lead dating evidence for an old western Grand Canyon on the assumption that groundwater table decline rates are equivalent to incision rates. Samples in the western Grand Canyon yielded apparent water table decline rates of 55 to 123 meters per million years over the past 17 million years, in contrast to eastern Grand Canyon samples that yielded much faster rates (166 to 411 meters per million years). Chronology and inferred incision data indicate that the Grand Canyon evolved via headward erosion from west to east, together with late-stage (~3.7 million years ago) accelerated incision in the eastern block.

Ever since the first geologist known to set eyes on the Grand Canyon, John Strong Newberry in 1858, and the famous John Wesley Powell expedition of 1869 (1), the age and origin of the Grand Canyon have remained a subject of great scientific and popular interest. Accurate incision rate data have, until now, come from dating basalt flows and travertine deposits, but these results have not been able to record both the downward and headward incision of the Grand Canyon over its entire history beyond 1 million years ago (Ma) and higher than 100 m above the river (2). More than 50 years ago, Arthur Lange, a speleologist, proposed that the study of cave sediments and speleothems (cave formations) could produce an accurate minimum age for the Grand Canyon (3). U-series dating of

speleothems, and consequently landscape evolution determinations using caves, began in the 1970s by alpha spectrometry (4) and were greatly

improved by the application of mass spectrometry in the mid-1980s (5).

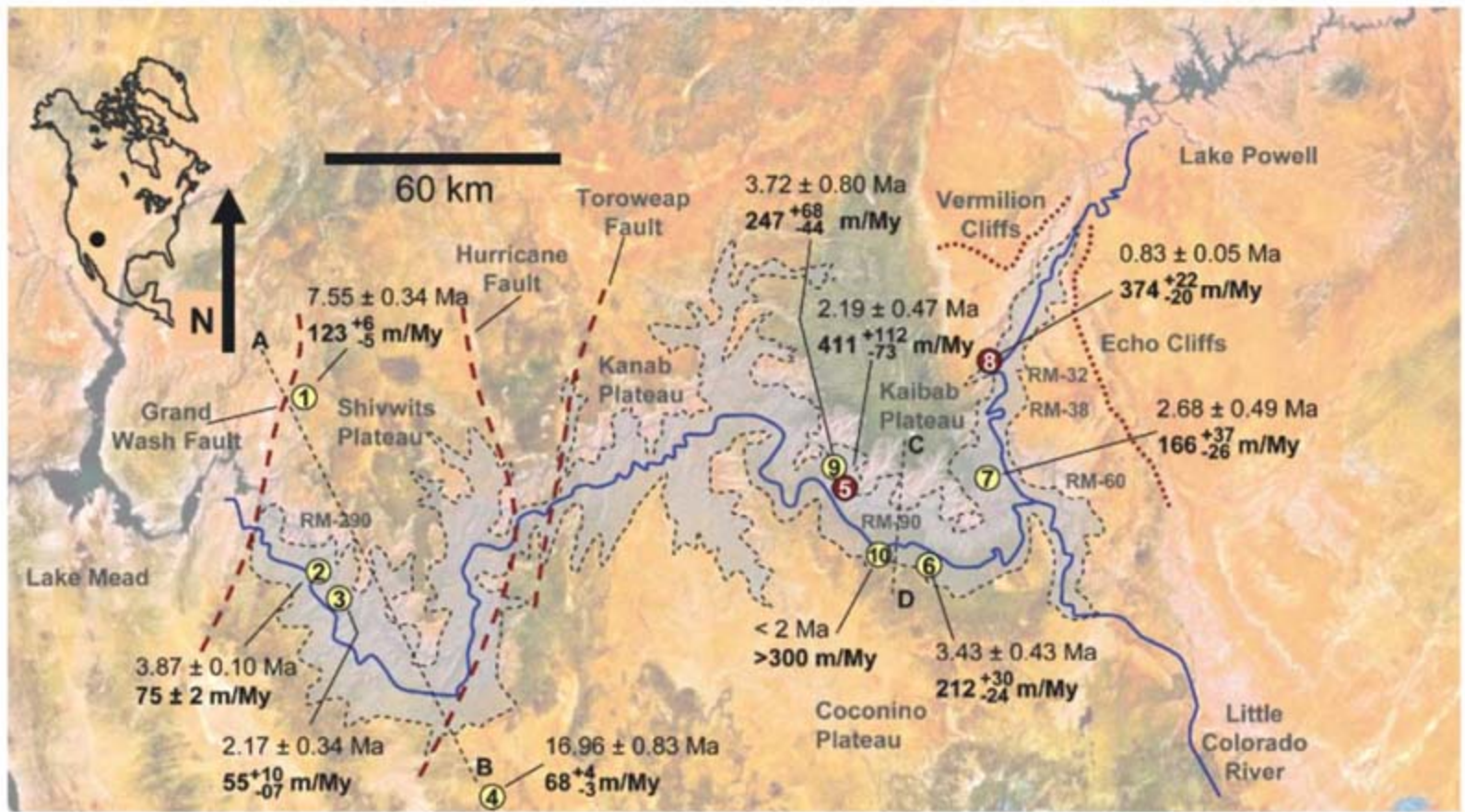
The realization that certain speleothems such as mammillary coatings (Fig. 1) form near groundwater tables [herein referred to as water tables (6)], and the fact that many Grand Canyon caves contain mammillary speleothems (7), has allowed us to take advantage of advances in U-Pb and U-series analytical techniques in an effort to make the long-sought chronology possible. For the Grand Canyon area (Fig. 2), there is no better niche than caves to find both clastic and chemical sediments that were deposited before, during, and after the incision of the canyon. Equally important, these cave deposits are located throughout the canyon. Caves are not only well suited to contain these deposits, they also provide an ideal environment that preserves and protects them from weathering. These mammillary coatings in the Grand Canyon caves contain sufficient uranium-lead ratios and yield U-Pb dates that place the water table within the canyon at a particular place and at an absolute time. This allows for the incision history of the Grand Canyon to be reconstructed



**Fig. 1.** Cave mammillaries coat cave walls below but near the water table. A cross section of broken mammillaries (m) and exposure of underlying bedrock (b) from site 6 (Tsean Bida) are shown. The unbroken form of this speleothem type (white arrows) indicates a subaqueous origin.

Department of Earth and Planetary Sciences, University of New Mexico, Albuquerque, NM 87131, USA.

\*To whom correspondence should be addressed. E-mail: [polyak@unm.edu](mailto:polyak@unm.edu)



**Fig. 2.** Map showing locations and U-Pb ages of cave mammillary samples and their apparent incision rates. Site numbers (in circles) are those referred to in Table 1 and the text; those in brown circles represent surface-exposed mammillary calcite. Washout satellite image was taken from the NASA World Wind Web site, with darker

regions representing higher elevations. Gray area is the canyon corridor. Two cross sections, A-B and C-D (fig. S3), show generalized pertinent stratigraphy. RM denotes the river-mile location. Incision rate errors assume  $\delta^{234}\text{U}_{\text{initial}}$  values = 3100‰ for sites 1, 2, 4, 6, and 9; see Fig. 3C for expanded uncertainties for these sites.

**Table 1.** U-Pb ages and incision rates from cave mammillaries. RM, river mile; IR, incision rate. Mother Cave mammillary age is estimated from  $\delta^{234}\text{U}_{\text{measured}}$  of  $17 \pm 3\text{‰}$  with  $\delta^{234}\text{U}_{\text{initial}}$  of  $3000 \pm 2500\text{‰}$ . For sites with  $\delta^{234}\text{U}_{\text{measured}} = 0\text{‰}$ ,  $\delta^{234}\text{U}_{\text{initial}}$  is assumed to be 3100‰. Extended  $2\sigma$  absolute errors on the incision rates assume a large uncertainty of the  $\delta^{234}\text{U}_{\text{initial}} = 3100 \pm 2500\text{‰}$ .

Site	Region	$^{238}\text{U}/^{206}\text{Pb}$ age (Ma)	$^{235}\text{U}/^{207}\text{Pb}$ age (Ma)	Concordia- constrained linear 3D age (Ma)	Dist. above river (m)	Dist. from river (km)	RM	IR (m/My)	$2\sigma$ error	Abs. error	Extended	
											$2\sigma$ error	Abs. error
1	Grand Wash Cliffs	$7.53 \pm 0.42$	$7.1 \pm 1.4$	$7.55 \pm 0.34$	930	38.6	277	123	+6	-5	+24	-18
2	Cave B	$3.8 \pm 0.32$	$4.3 \pm 0.5$	$3.87 \pm 0.10$	290	0.5	266	75	+2	-2	+35	-15
3	Dry Canyon	$2.17 \pm 0.42$	$8.1 \pm 9.9$	$2.17 \pm 0.34$	120	1.6	265	55	+10	-7		
4	Grand Canyon Caverns	$17.3 \pm 1.60$	$29.0 \pm 14.0$	$16.96 \pm 0.83$	1160	28.9	190	68	+4	-3	+4	-3
5	Gavain Abyss	$2.39 \pm 0.77$	$6.2 \pm 5.9$	$2.19 \pm 0.47$	900	5.5	93	411	+112	-73		
6	Tsean Bida	$3.37 \pm 0.50$	$1.0 \pm 16.0$	$3.43 \pm 0.43$	726	4.6	80	212	+30	-24	+134	-59
7	Butte Fault Cave	$2.73 \pm 0.63$	$3.7 \pm 7.9$	$2.68 \pm 0.49$	445	2.6	57	166	+37	-26		
8	Bedrock Canyon	$0.8 \pm 0.12$	$0.7 \pm 0.3$	$0.83 \pm 0.05$	310	2.1	32	374	+22	-20		
9	Shinumo Creek Cave	$3.5 \pm 1.30$	$-1.0 \pm 5.2$	$3.72 \pm 0.80$	920	6.6	94	247	+68	-44	+208	-78
10	Mother Cave		$^{234}\text{U}$ age = $1.6 \pm 0.5$		605	2.2	90	>300				

on the basis of the premise that the timing of this water table descent is coeval with incision of the Colorado Plateau by the Colorado River system.

There are three main reasons for placing the deposition of mammillary coatings at or near the water table. First, the fine-grained, fibrous nature of these speleothems attests to the rapid degassing of  $\text{CO}_2$  and precipitation of calcite near the water table surface. Second, the Grand Canyon mammillaries are often associated with folia and cave rafts, two speleothem types known to form directly at the water table (6). Third, mammillaries can be seen forming near the water table today in caves such as Devils Hole, Nevada (380 km west-northwest of our study area), where they occur along the top few tens of meters of the water table (8, 9). The presence of gypsum rinds in Grand Canyon caves—interpreted to be speleogenetic-type crusts formed by oxidation of  $\text{H}_2\text{S}$  diffusing upward into the caves as the water table drops through the caves (fig. S1)—also supports mammillary association with the water table. These gypsum rinds, which directly overlie mammillary coatings, form just above the water table and display characteristic sulfur isotope ( $\delta^{34}\text{S}$ ) values depending on the source of the  $\text{H}_2\text{S}$  (10).

Mammillary samples were dated by the U-Pb method (11). U-Pb analyses of relatively young carbonate speleothems have recently been shown

to be feasible under certain circumstances (12), and concerns with excess  $^{206}\text{Pb}$  from initial  $^{234}\text{U}$ - $^{238}\text{U}$  isotopic disequilibrium can be resolved by combining U-Pb data with  $^{234}\text{U}$ - $^{238}\text{U}$  chronometer data (12, 13):  $^{206}\text{Pb}$  data are corrected for  $^{234}\text{U}$  excess and  $^{230}\text{Th}$  deficiency, and  $^{207}\text{Pb}$  data are corrected for  $^{231}\text{Pa}$  deficiency (9) (fig. S2). Where U-Pb and  $^{234}\text{U}$  chronologies overlap, the two systems give consistent chronology.

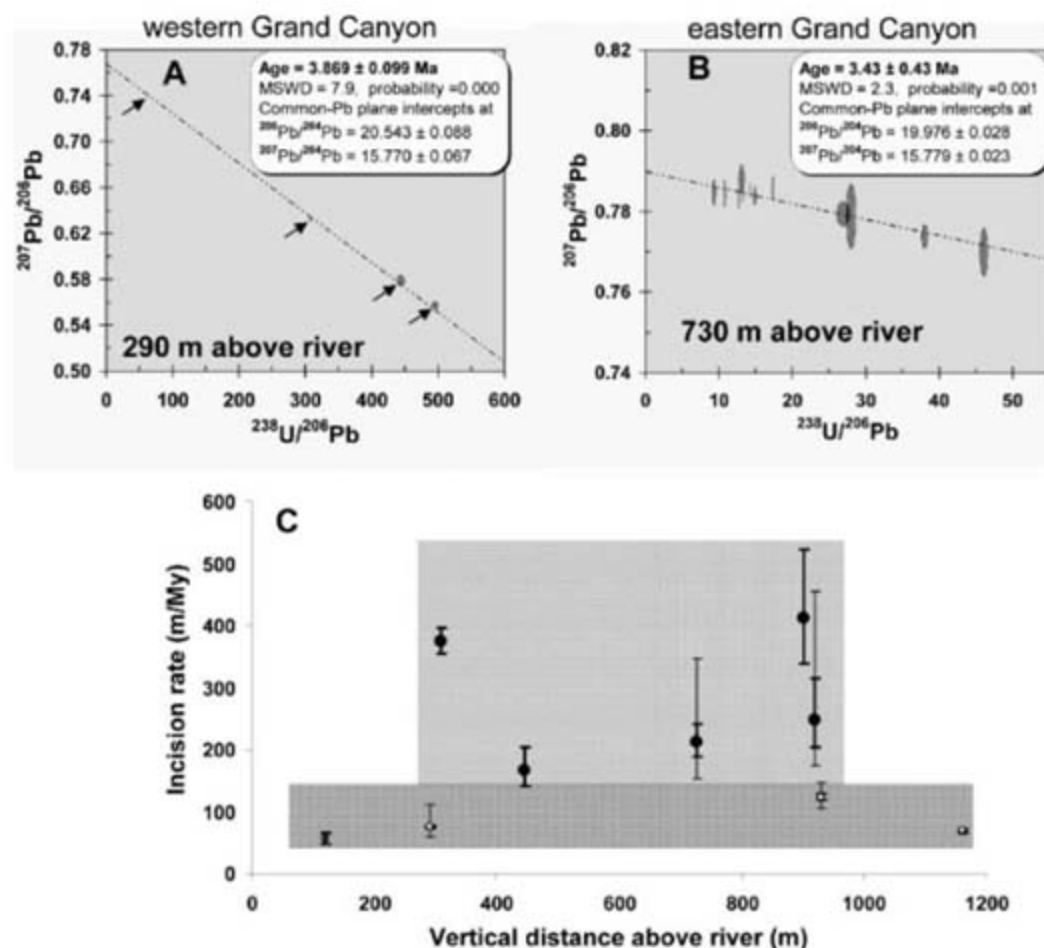
Our core data (57 analyses) come from nine mammillary coatings throughout the canyon, referred to as sites 1 to 9 (Fig. 2). Four of these coatings (sites 1 to 4) for the western Grand Canyon, all within 1200 vertical meters above the Colorado River, yield ages of 17.0, 7.6, 3.9, and 2.2 Ma. Five other coatings (sites 5 to 9) are located in the eastern Grand Canyon, all within 950 vertical meters above the river, and have ages of 3.7, 3.4, 2.7, 2.2, and 0.8 Ma. For simplicity and consistency, all apparent water table descent rates are based on a relatively flat water table over time. Our U-Pb ages (Table 1 and table S1) show water table descent rates of 55 to 123 m per million years (My) in the western Grand Canyon and 166 to 411 m/My in the eastern Grand Canyon.

Sample sites 1 and 4 are situated 42 and 26 km north and south, respectively, of the Colorado River (Fig. 2) and show water table descent rates that have spatial extent just beyond the canyon

itself. Our four western data points support a relatively stable slow drop in the water table in the western Grand Canyon over the past 17 My. In contrast, results from the eastern Grand Canyon show distinctly faster water table descent rates, all having U-Pb ages less than 4 Ma (Fig. 3). In addition, one other site in the eastern Grand Canyon could not be dated by U-Pb (excess common Pb) but has a positive  $\delta^{234}\text{U}$  value. This sample from Mother Cave near Grand Canyon Village (site 10; Fig. 2 and Table 1) has a  $^{234}\text{U}$ - $^{238}\text{U}$  age [based on a  $\delta^{234}\text{U}$  initial value of  $3000 \pm 2500$  per mil (‰)] of  $1.6 \pm 0.5$ , yielding an apparent water table descent rate of  $>300$  m/My; these data provide further evidence of faster eastern Grand Canyon water table descent rates.

Western Grand Canyon incision rates are well constrained at 50 to 75 m/My for the last 0.73 My from basalts located within 60 vertical meters above the river (2). Our incision rates representing 100 to 1200 m of incision (55 to 75 m/My, excepting site 1) compare well with those determined from the near-river basalt flows and fill an incision rate history gap representing millions of years. Canyon incision by a smaller river system is the most likely interpretation for the relatively stable long-lasting rate of apparent water table descent (55 to 123 m/My) for the western Grand Canyon. This suggests that the western Grand Canyon has been forming for the past 17 My and has probably progressed slowly headward to the east over that entire period. An older western Grand Canyon fits nicely with Miocene extension and timing of Grand Wash fault activity that may represent uplift of the western edge of the Colorado Plateau just prior to 20 Ma (14) and filling of the Grand Wash trough with sediment coming from the Virgin Mountains to the north and higher topography just to the east (15). A water table descent to the elevation of site 1 (~1200 m above sea level) indicates that erosion of trough sediments was taking place as early as 7.6 Ma. Our incision rate data would imply that Grand Canyon incision into the Grand Wash fault cliffs incised into the top of the Redwall-Muav aquifer of the Colorado Plateau some time between 16 and 9 Ma. Initiation of deposition of the Miocene-aged, lacustrine, Hualapai limestone [11 to 6 Ma (16, 17)] may have been coeval with, and the result of, the release of carbonate-rich water from the newly truncated Redwall Limestone (14, 18).

The eastern third of the Grand Canyon appears to have undergone fast incision ( $>166$  m/My) and rapid headward erosion starting before 3.7 Ma and likely at 5 to 6 Ma when the Colorado River became fully integrated and through-flowing (17). A previously reported faster incision rate in the eastern Grand Canyon [150 to 230 m/My for the past 0.5 My; river mile 60 (19)] is consistent with our results in that area (140 to 203 m/My) and was compared and attributed to the Hurricane Fault displacement (2). However, our data suggest accelerated headward erosion in the eastern Grand Canyon—whether from simple



**Fig. 3.** (A and B) U-Pb Concordia-constrained linear three-dimensional isochron ages for samples in the western Grand Canyon (A) and the eastern Grand Canyon (B). Note the difference in elevations of these two samples of similar age. These data support a headward erosion scenario for the Grand Canyon. (C) Graph illustrating the distinct differences in incision rates in the western versus eastern Grand Canyon. Extended error bars assume a large uncertainty of the  $\delta^{234}\text{U}_{\text{initial}}$  values =  $3100 \pm 2500$ ‰ for sites 1, 2, 4, 6, and 9.



knick-point propagation, lake overflow (20), or karst capture (18). Marble Canyon mammillaries also indicate fast incision rates. At river miles 57 (site 7) and 32 (site 8), incision rates are 166 and 374 m/My, respectively, and at river mile 38, a speleogenetic gypsum crust yielded a U-series age of 0.15 Ma at 50 m above the river and an incision rate of  $\leq 330$  m/My.

Even though our interpretation assumes no structural or hydrologic complexities, those complexities could help to explain any apparent inconsistencies as data accumulate. However, it is likely that the structure and hydrology will not change the overall interpretation that the western Grand Canyon is older than the eastern Grand Canyon. For instance, an explanation for the higher incision rate at site 1 (Bobcat Cave, Grand Wash Cliffs) could be the presence of an elevated water table north of the western Grand Canyon at 7.6 Ma. Even without this interpretation, the highest western Grand Canyon incision rate from the Grand Wash Cliffs mammillary of 123 m/My (site 1) is still less than the slowest incision rate for the eastern Grand Canyon (166 m/My, site 7; Fig. 3 and fig. S3C). The fast incision rate at Bedrock Canyon (site 8) indicates that 300 m of incision occurred in Marble Canyon over the past 1 My. However, at river miles 57 (site 7) and 60 (19), incision rates are slower ( $\sim 140$  to 230 m/My). Structure, hydrology, or headward erosion history (i.e., knick-point propagation from site 7 to site 8) might resolve these differences when additional data are available.

Overall, our data argue for an older Grand Canyon that was modified in the late Miocene by

a fully integrated Colorado River that accelerated the headward erosion of the eastern Grand Canyon. We found that mammillary calcite is not restricted to large caves in the Redwall and Muav limestones, but also occurs in small fissure-controlled caves in other units such as the overlying Supai Formation. Hundreds of these deposits probably exist throughout the canyon, offering the potential for a reconstruction of the canyon's history, with a resolution perhaps high enough to explain complexities of the canyon's history related to faults, folds, and volcanic and tectonic activity.

#### References and Notes

1. W. Ranney, *Carving Grand Canyon—Evidence, Theories, and Mystery* (Grand Canyon Association, Grand Canyon, AZ, 2005).
2. K. E. Karlstrom et al., *Geol. Soc. Am. Bull.* **119**, 1283 (2007).
3. A. L. Lange, *Plateau* **27**, 1 (1955).
4. T. C. Atkinson, P. J. Rowe, in *Uranium-Series Disequilibrium: Applications to Earth, Marine, and Environmental Sciences*, M. Ivanovich, R. Harmon, Eds. (Oxford Univ. Press, New York, 1992), pp. 669–703.
5. R. L. Edwards, J. H. Chen, G. J. Wasserburg, *Earth Planet. Sci. Lett.* **81**, 175 (1986).
6. C. A. Hill, P. Forti, *Cave Minerals of the World* (National Speleological Society, Huntsville, AL, 1997).
7. V. J. Polyak, C. A. Hill, Y. Asmeron, *GSA Abs. Prog.* **36**, 549 (2004).
8. I. J. Winograd et al., *Science* **242**, 1275 (1992).
9. B. J. Szabo, P. T. Kolisar, A. C. Riggs, I. J. Winograd, K. R. Ludwig, *Quat. Res.* **41**, 56 (1994).
10. C. A. Hill, V. J. Polyak, W. C. McIntosh, P. P. Provencio, in *Colorado River Origin and Evolution*, R. A. Young, E. E. Spamer, Eds. (Grand Canyon Association, Grand Canyon, AZ, 2001), pp. 141–146.
11. See supporting material on Science Online.
12. D. A. Richards, S. H. Bottrell, R. A. Cliff, K. Strohle, P. J. Rowe, *Geochim. Cosmochim. Acta* **62**, 3683 (1998).

13. J. Woodhead et al., *Quat. Geochronol.* **1**, 208 (2006).
14. J. E. Faulds, L. M. Price, M. A. Wallace, in *Colorado River Origin and Evolution*, R. A. Young, E. E. Spamer, Eds. (Grand Canyon Association, Grand Canyon, AZ, 2001), pp. 93–100.
15. L. S. Beard, *GSA Spec. Pap.* **303**, 27 (1996).
16. J. E. Faulds, M. A. Wallace, L. A. Gonzalez, M. T. Heizler, in *Colorado River Origin and Evolution*, R. A. Young, E. E. Spamer, Eds. (Grand Canyon Association, Grand Canyon, AZ, 2001), pp. 81–87.
17. J. E. Spencer, L. Peters, W. C. McIntosh, P. J. Patchett, in *Colorado River Origin and Evolution*, R. A. Young, E. E. Spamer, Eds. (Grand Canyon Association, Grand Canyon, AZ, 2001), pp. 89–92.
18. C. A. Hill, N. Eberz, R. H. Buecher, *Geomorphology* **95**, 316 (2008).
19. J. Pederson, K. Karlstrom, W. Sharp, W. McIntosh, *Geology* **31**, e17 (2003).
20. N. Meek, J. Douglass, in *Colorado River Origin and Evolution*, R. A. Young, E. E. Spamer, Eds. (Grand Canyon Association, Grand Canyon, AZ, 2001), pp. 199–206.
21. Supported by NSF grant EAR-0518602 and by a National Park Service research grant. We thank E. Benenati and K. Voyles for the necessary collection permits; the Navajo, Hopi, and Hualapai tribes for permission to enter and collect from caves on their land; P. Provencio, A. Hill, C. Mosch, and M. Goar for field assistance; S. Davis, D. Powell, M. Oliphant, and N. Pistole for helping us to gain access to remote areas along the river; A. and P. Palmer for help in the field and through cave science discussions along the river; and especially B. and D. Buecher and D. Powell for their tremendous amount of fieldwork and sample elevation-location data.

#### Supporting Online Material

[www.sciencemag.org/cgi/content/full/319/5868/1377/DC1](http://www.sciencemag.org/cgi/content/full/319/5868/1377/DC1)

Materials and Methods

Figs. S1 to S3

Table S1

References

2 October 2007; accepted 30 January 2008

10.1126/science.1151248

## The Dust Halo of Saturn's Largest Icy Moon, Rhea

G. H. Jones,<sup>1,2,3\*</sup> E. Roussos,<sup>1</sup> N. Krupp,<sup>1</sup> U. Beckmann,<sup>4</sup> A. J. Coates,<sup>2,3</sup> F. Crary,<sup>5</sup> I. Dandouras,<sup>6</sup> V. Dikarev,<sup>1,4,7</sup> M. K. Dougherty,<sup>8</sup> P. Garnier,<sup>6,9</sup> C. J. Hansen,<sup>10</sup> A. R. Hendrix,<sup>10</sup> G. B. Hospodarsky,<sup>11</sup> R. E. Johnson,<sup>12</sup> S. Kempf,<sup>4</sup> K. K. Khurana,<sup>13</sup> S. M. Krimigis,<sup>14,15</sup> H. Krüger,<sup>1</sup> W. S. Kurth,<sup>11</sup> A. Lagg,<sup>1</sup> H. J. McAndrews,<sup>2,19</sup> D. G. Mitchell,<sup>14</sup> C. Paranicas,<sup>14</sup> F. Postberg,<sup>4</sup> C. T. Russell,<sup>13</sup> J. Saur,<sup>16</sup> M. Seiß,<sup>17</sup> F. Spahn,<sup>17</sup> R. Srama,<sup>4</sup> D. F. Strobel,<sup>18</sup> R. Tokar,<sup>19</sup> J.-E. Wahlund,<sup>9</sup> R. J. Wilson,<sup>19</sup> J. Woch,<sup>1</sup> D. Young<sup>5</sup>

Saturn's moon Rhea had been considered massive enough to retain a thin, externally generated atmosphere capable of locally affecting Saturn's magnetosphere. The Cassini spacecraft's in situ observations reveal that energetic electrons are depleted in the moon's vicinity. The absence of a substantial exosphere implies that Rhea's magnetospheric interaction region, rather than being exclusively induced by sputtered gas and its products, likely contains solid material that can absorb magnetospheric particles. Combined observations from several instruments suggest that this material is in the form of grains and boulders up to several decimetres in size and orbits Rhea as an equatorial debris disk. Within this disk may reside denser, discrete rings or arcs of material.

On 26 November 2005, Cassini encountered Rhea, the second largest of Saturn's moons, at 500 km altitude, detecting in situ the anticipated (1, 2), approximately spherical distribution of grains lofted from its surface

by interplanetary dust impacts. Cassini passed downstream of Rhea with respect to the local magnetospheric flow (Fig. 1) and observed the anticipated wake caused by plasma striking the moon, together with an unpredicted depletion of

energetic electrons extending to  $\sim 8$  Rhea radii ( $R_R$ ) (Fig. 2). The scale of the depletion indicates that some material is absorbing electrons within the volume dominated by Rhea's gravitational field: its Hill sphere, of radius 7.7  $R_R$ . Voyager 1 measurements in 1980 previously indicated a broadened depletion's presence farther downstream (3). A more distant Cassini flyby in August 2007 also showed evidence of a broad electron depletion (4). No such features have yet been observed at Dione and Tethys, but energetic electrons are absorbed by grains ejected from Enceladus's south pole (5). The signature's profile indicates that the electron-absorbing material has a near-symmetrical distribution about Rhea. No evidence was found by Cassini's instruments for the presence of large amounts of freshly ionized gas, which could theoretically scatter electrons. Neutral gas and dust populations are therefore the primary absorbing-medium candidates.

As Cassini passed Rhea, its cosmic dust analyzer (CDA) (6) registered an increase in the impact rate of  $>1$   $\mu\text{m}$  particles (Fig. 3), signaling the predicted envelopment of the moon (1, 2) by dust ejected by micrometeoroid impacts. An impact-ejecta model (1) indicates a

preliminary interplanetary projectile flux of  $5 \times 10^{-15} \text{ kg m}^{-2} \text{ s}^{-1}$  (4), liberating  $\sim 230 \text{ kg s}^{-1}$  of dust, corresponding to  $7.2 \times 10^{14}$  particles  $\text{s}^{-1}$  larger than  $1 \mu\text{m}$ . CDA and radio and plasma wave science instrument (RPWS) detected a dust distribution sharply peaked near closest approach. The magnetospheric imaging instrument (MIMI) ion and neutral camera (INCA) recorded a flatter distribution extending beyond the Hill sphere, suggesting that it is most sensitive to small grains, and consequently that smaller Rhea impact ejecta have higher velocities. Unlike the electron signatures, there is no clear dust signal coincident with the Hill sphere boundary: The electron fluxes reflect the integrated effects of electrons' traverses of an absorbing medium before reaching Cassini's energetic electron detectors; most absorbing material is inferred to reside north of Cassini's path (4). Fluxes were slightly enhanced outbound in RPWS and MIMI-INCA data, when Cassini was nearest Rhea's equatorial plane.

Neutral gas surrounding Rhea is liberated from the moon and its dust halo through dust impacts and magnetospheric particle sputtering (7). Upper limits in column density deduced using the ultraviolet imaging spectrograph (UVIS) (8) are  $1.5 \times 10^{13} \text{ cm}^{-2}$  for O and  $1.6 \times 10^{14} \text{ cm}^{-2}$  for  $\text{O}_2$ . Independently derived MIMI-INCA upper limits (fig. S1) are sensitive to the assumed background ion flux; the upper limits for the same species are in the range of  $10^{13}$  to  $10^{14} \text{ cm}^{-2}$  (see SOM text).

The quantity of electron-absorbing material can be estimated by gauging electron path

lengths within the Hill sphere. Most  $<700 \text{ keV}$  electrons observed by MIMI and the Cassini plasma spectrometer (CAPS) traverse the Hill sphere several times north-south before detection, bouncing along Saturn's convecting magnetic field lines (9). The reduced fluxes reflect the integrated effects of gas and dust upstream of Cassini's location  $\sim 200 \text{ km}$  south of Rhea's equatorial plane. We modeled the effect of an electron-absorbing, spherically symmetric gas cloud required to produce the broad electron depletion. Implied neutral gas column densities of  $\sim 10^{18} \text{ cm}^{-2}$  near Rhea's surface easily exceeded the column density upper limits. We next modeled a spherically symmetric dust halo, the densest possible based on CDA data, with number density falling off as an inverse power law with radial distance of  $-2.5$  and a grain size of  $\sim 10 \mu\text{m}$ . Electron number densities were assumed to decrease by a factor of  $e^2$  when the column mass encountered equaled the electrons' range in water (10, 11). Coulomb collisions were excluded. There was no unique solution, but surface densities required were  $>2$  orders of magnitude higher than implied by CDA data. The  $-2.5$  density falloff is too steep to explain the detectable depletion at the Hill sphere's edge.

To explain the observations, the presence of an additional electron-absorbing obstacle is implied. An analysis of the electron data indicates that this obstacle is most likely in the form of a low optical depth disk of material near Rhea's equatorial plane and that the disk contains solid bodies up to  $\sim 1 \text{ m}$  in size (4). Dynamical studies show that such bodies could orbit persistently near the equatorial plane. The disk explains the electron signature and is consistent with in situ data south of the equatorial plane, because Cassini did not traverse the plane itself. There have been no reports of remote observations of such a disk, but it would not be easy to detect (4). A

weakly perceptible population of neutral gas and related products could also be present. Neutral gas is primarily lost through ionization and subsequent transport by the magnetospheric plasma. Several plasma measurements imply that there is a detectable, although not fully understood, interaction between Saturn's magnetosphere and freshly ionized material around Rhea, at the scales of the moon's Hill sphere (Fig. 2). As Cassini traversed the Hill sphere, electron number densities increased slightly at  $\sim 22:25 \text{ UT}$ , near the inbound boundary, but a similar outbound feature is difficult to identify. Bulk velocities derived from the thermal plasma ion data shown in Fig. 2E were below corotation within the Hill sphere and displayed a  $\leq 10^\circ$  anti-Saturnward flow deviation inbound.

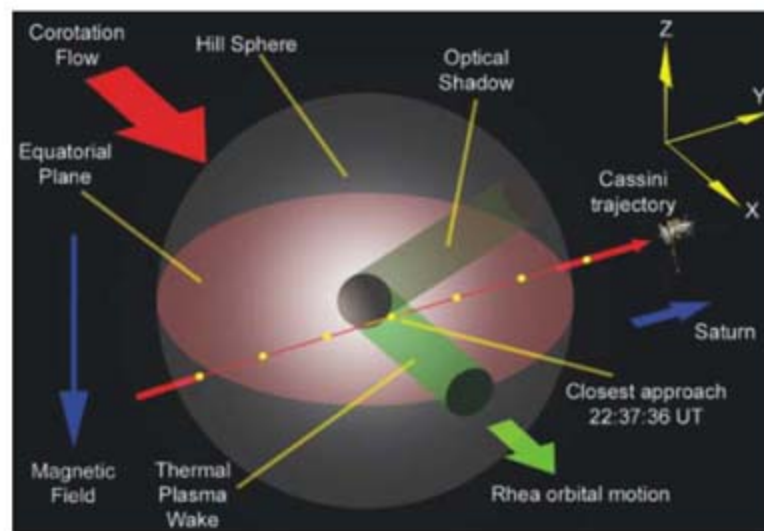
Several short-duration dips in MIMI low-energy magnetospheric measurement system (LEMMS) electron flux are also evident, lasting a few seconds each, or tens of km along the spacecraft track, and are nearly symmetrical about Rhea (Fig. 4). No plasma-related cause for these narrow electron dips has been identified. They are not believed to be electron phase-space holes, which usually result from a two-stream instability, as may occur in the wakes of absorbing bodies (12). Such structures are typically traversed in tens of milliseconds as opposed to several seconds. The dips are, however, consistent with additional obstacles residing upstream or directly north of Cassini, in Rhea's equatorial plane, that cause the occultation of electrons before detection at Cassini. The broadening of the dips with increasing electron energy (fig. S8) is as expected for an obstacle to electrons; the widening reflects electrons' larger gyroradii and slower convection through the Hill sphere at higher energies, and hence more prolonged absorption. The simplest explanation of these near-symmetrical features is that extended arcs or rings of material orbit Rhea

<sup>1</sup>Max-Planck-Institut für Sonnensystemforschung, Max-Planck-Strasse 2, 37191 Katlenburg-Lindau, Germany. <sup>2</sup>Mullard Space Science Laboratory, Department of Space and Climate Physics, University College London, Holmbury St. Mary, Dorking, Surrey RH5 6NT, UK. <sup>3</sup>Centre for Planetary Sciences, University College London, London WC1E 6BT, UK. <sup>4</sup>Max Planck Institut für Kernphysik, Saupfercheckweg 1, 69117 Heidelberg, Germany. <sup>5</sup>Southwest Research Institute, Culebra Road, San Antonio, TX 78238, USA. <sup>6</sup>Centre d'Etude Spatiale des Rayonnements, Paul Sabatier University, Centre National de la Recherche Scientifique, 9 Avenue du Colonel Roche, Toulouse, 31400 France. <sup>7</sup>Astronomical Institute of St. Petersburg State University, Universitetskii pr. 28, Staryi Peterhof, St. Petersburg, 198504, Russia. <sup>8</sup>The Blackett Laboratory, Imperial College London, London SW7 2BW, UK. <sup>9</sup>Swedish Institute of Space Physics, Box 537, 751 21 Uppsala, Sweden. <sup>10</sup>Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA. <sup>11</sup>Department of Physics and Astronomy, The University of Iowa, Iowa City, Iowa 52242, USA. <sup>12</sup>Department of Materials Science and Engineering, University of Virginia, Post Office Box 400745, 116 Engineer's Way, Charlottesville, VA 22904, USA. <sup>13</sup>Institute of Geophysics and Planetary Physics, University of California, Los Angeles, 6863 Slichter Hall, Los Angeles, CA 90095, USA. <sup>14</sup>The Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, MD 20723, USA. <sup>15</sup>Academy of Athens, Soranou Efessiou 4, Athens 115 27, Greece. <sup>16</sup>Institut für Geophysik und Meteorologie, Universität zu Köln, Albertus-Magnus-Platz, 50923 Köln, Germany. <sup>17</sup>AG Nichtlineare Dynamik, Universität Potsdam, Postfach 601553, 14469 Potsdam, Germany. <sup>18</sup>Department of Earth and Planetary Science, The Johns Hopkins University, Baltimore, MD 21218, USA. <sup>19</sup>Los Alamos National Laboratory, Los Alamos, NM 87545, USA.

\*To whom correspondence should be addressed. E-mail: ghj@mssl.ucl.ac.uk

**Fig. 1.** Geometry of the 2005 flyby. Rhea is continuously overtaken at  $\sim 58 \text{ km s}^{-1}$  by cold Saturnian magnetospheric plasma (7) flowing in the direction of the moon's orbital motion. Cassini's path through Rhea's Hill sphere was downstream of the moon with respect to this flow, entering at 22:24:39 when 290 km south of the equatorial plane and leaving at 22:50:42 UT when 160 km south.

Spacecraft orientation was fixed throughout. Closest approach to the 1529-km-diameter moon's leading hemisphere occurred 228 km south of the equatorial plane, when Cassini was 1244 km downstream of Rhea's center. Points along the trajectory are separated by 5 min. The Cartesian coordinate system shown, displaced here and in other figures for clarity from its origin at Rhea's center, has components along Rhea's orbital motion vector (X), Saturnward (Y), and perpendicular to Rhea's orbital plane (Z).



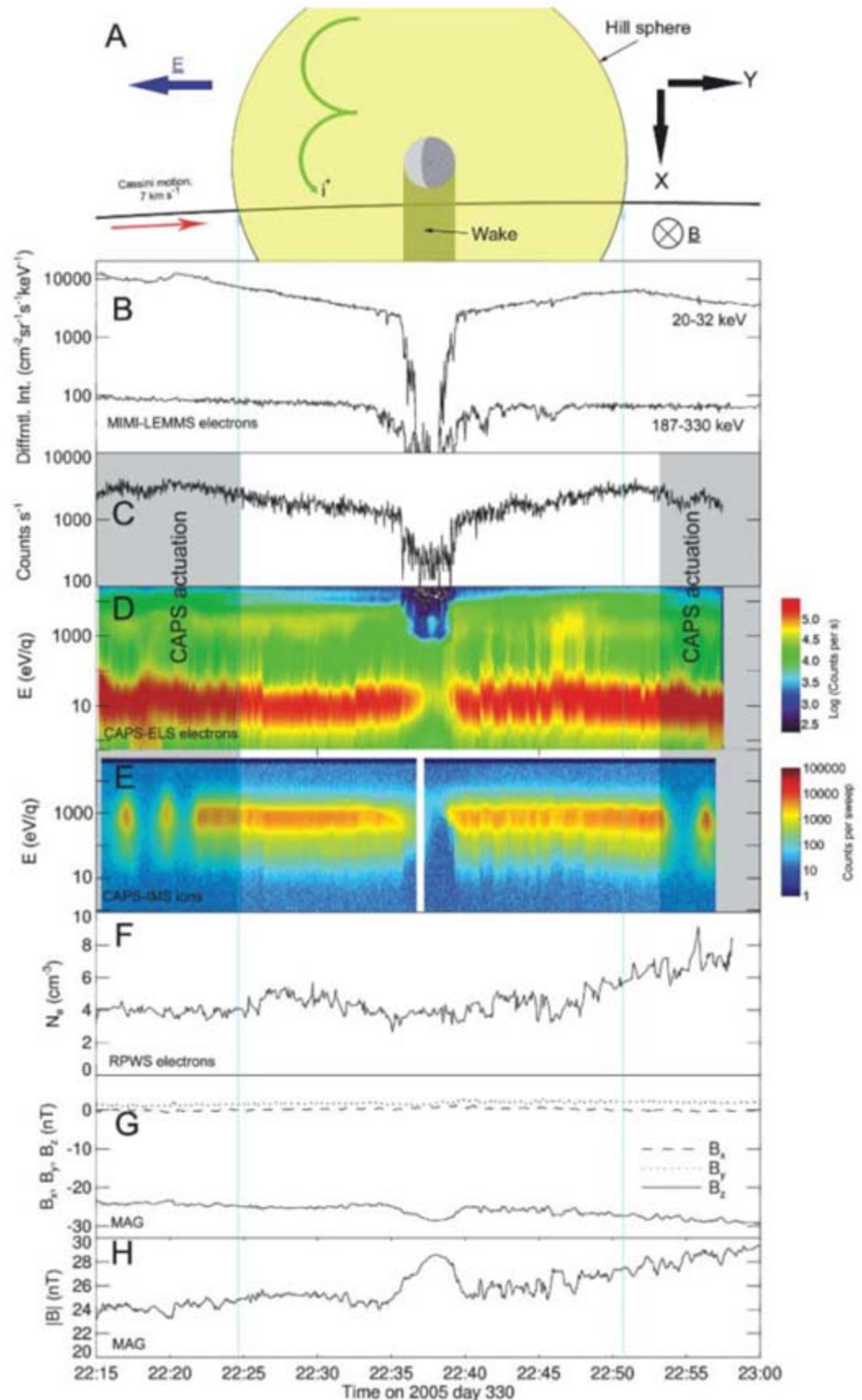
in its equatorial plane, with orbital radii of ~1610, ~1800, and ~2020 km. The slight asymmetry in dip positions may be caused by a modest tilt to the local magnetic field.

The dips may instead represent the cumulative effect of material upstream of Cassini: Saturn's corotational plasma flows across the rings, rather than parallel to them as occurs at Saturn itself. The dips may therefore occur

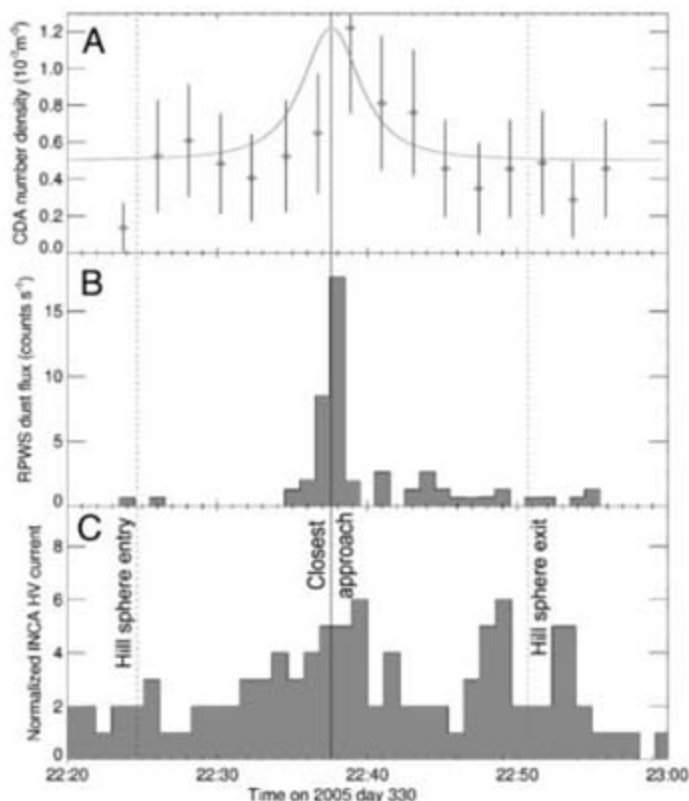
downstream of local maxima in ring column density presented to the corotating plasma, that is, the ansae, where narrow rings' tangents directly upstream of Cassini are parallel to the flow. For this scenario, assuming perfectly circular rings, common plasma flow deviations (5) could explain the slight asymmetry in the dips' locations. A further complication of both ring interpretations is the observed electrons' multiple traverses of

the Hill sphere before detection. If an integer number of electron half bounce-lengths separates Cassini from a single ring, multiple "ghost" dips may form through occasional constructive interference of weaker depletions. Dust lofted by plasma-induced charging of Rhea's regolith (13) would have been a candidate nonring obstacle, especially as near-zero grain potentials at Rhea (14) may make its surface particularly sensitive to

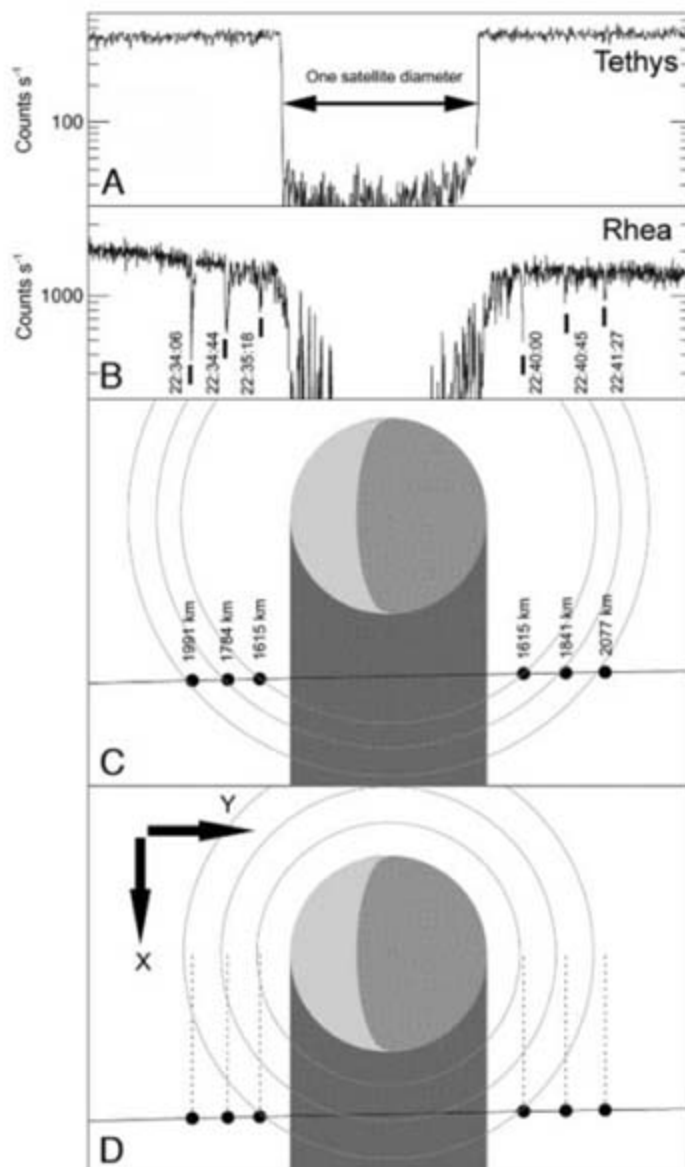
**Fig. 2.** Data returned during the flyby. **(A)** Rhea and its Hill sphere to the same scale as the other panels. The general directions of the local magnetic **[B]** and corotational electric **[E]** fields are shown, together with the initial gyromotion of a freshly created pickup ion,  $i^+$ . Prominent in almost all data sets is Rhea's infilling wake (22:35:30 to 22:39:40 UT). **(B)** MIMI (19) LEMMS 20 to 32 keV and 187 to 330 keV. keV electron fluxes at pitch angles  $\sim 10^\circ$ . **(C)** CAPS (20), electron spectrometer (ELS) 26 keV electron fluxes at pitch angles of  $\sim 90^\circ$ . **(D)** CAPS-ELS electron fluxes. **(E)** CAPS ion mass spectrometer (IMS) ion spectrogram (all anodes), showing gyro-radius effects in Rhea's wake. **(F)** Total electron number density from RPWS (21) upper hybrid frequency values. **(G)** Components and **(H)** magnetic field magnitude from Cassini's magnetometer (22, 23). Magnetic field variations appear consistent with Rhea being plasma-absorbing: The evacuated wake's field strength is enhanced to maintain constant total particle plus magnetic pressure. The magnetic field wave power decreases somewhat within the Hill sphere, probably due to charged particle loss causing decreased plasma  $\beta$ ; otherwise, no direct magnetic field perturbations are convincingly evident on the Hill sphere's scale. Cassini's flyby trajectory, almost confined to the interaction region's equatorial symmetry plane, where field rotations are minimized, and almost transverse to the north-south symmetry plane, makes a decisive interpretation of the magnetic signature difficult.



**Fig. 3.** In situ dust observations. (A) Comparison of recalibrated and dead time-corrected  $>1\text{-}\mu\text{m}$  CDA data to a modeled isotropic impactor-ejecta number density (1). Two other instruments provide in situ dust data: RPWS (21) detects expanding plasma clouds resulting from high-velocity dust impacts onto Cassini (24), while MIMI-INCA (19) records dust-induced discharges between its collimating plates. (B) The RPWS dust flux, estimated to respond to particles  $>3\text{-}\mu\text{m}$ . (C) MIMI-INCA high-voltage current. Although ultraviolet light may cause spurious signals in the INCA data set, dust impacts are believed to dominate.



**Fig. 4.** Interpretation of electron flux dropouts as evidence of a ring system. (A) 28 to 49 keV electron fluxes at Tethys, demonstrating that moon's sharp, isolated absorption signature, in contrast to equivalent data from Rhea (B). The panels are scaled such that the relative size of the satellites' diameters is equal. Rhea's wake infills more rapidly because of the higher local plasma temperature. (C) Locations of some short-lived electron flux dropouts visible in the MIMI-LEMMS electron flux shown in (B). Their distances from Rhea's rotational axis are shown. Such absorptions, reminiscent of other planetary ring signatures (25), could be explained by the occultation of southward-traveling electrons before detection by Cassini. The signatures' spatial near-symmetry may signal the presence of three near-circular rings or arcs of material  $\sim 230\text{ km}$  north, in Rhea's equatorial plane. (D) An alternative dropout formation scenario.



such processes. However, several discrete sheets of grains as implied by multiple dips cannot be easily explained.

Numerical simulations imply that bodies that are subject only to minor perturbations by non-gravitational forces could orbit Rhea stably for long periods (15). Impacts on Rhea large enough to result in the orbit of ejecta material may have occurred as recently as 70 million years ago (16). Disruptive capture of a minor body is also possible. In these cases, some debris could have remained in orbits that flattened and circularized over time. Clumps of Rhea-orbiting material could maintain this disk, and bombardment by E-ring debris and extra-Saturnian grains would also release ejecta into near-circular orbits. Isolated clumps or moonlets could also explain the material's gravitational containment in discrete rings or arcs. Given their orbital stability, rings may even have survived since Rhea's formation.

#### References and Notes

1. A. V. Krivov *et al.*, *Planet. Space Sci.* **51**, 251 (2003).
2. M. Sremcevic, A. V. Krivov, F. Spahn, *Planet. Space Sci.* **51**, 455 (2003).
3. J. F. Carbary, S. M. Krimigis, W.-H. Ip, *J. Geophys. Res.* **88**, 8947 (1983).
4. Materials, methods, and further discussion are available as supporting material on Science Online.
5. G. H. Jones *et al.*, *Science* **311**, 1412 (2006).
6. R. Srama *et al.*, *Space Sci. Rev.* **114**, 465 (2004).
7. J. Saur, D. F. Strobel, *Astrophys. J.* **620**, L115 (2005).
8. L. W. Esposito *et al.*, *Space Sci. Rev.* **115**, 299 (2004).
9. M. F. Thomsen, J. A. Van Allen, *J. Geophys. Res.* **85**, (A11), 5831 (1980).
10. National Institute of Standards and Technology Physical Reference Data, <http://physics.nist.gov/PhysRefData/>, (2006).
11. J. N. Cuzzi, J. A. Burns, *Icarus* **74**, 284 (1988).
12. P. C. Birch, S. Chapman, *Phys. Plasmas* **9**, 1785 (2002).
13. T. J. Stubbs *et al.*, abstr. 1899, *Lunar and Planetary Science Conference XXXVI*, Lunar and Planetary Institute (League City, TX, 14 to 18 March 2005).
14. S. Kempf *et al.*, *Planet. Space Sci.* **54**, 999 (2006).
15. Test particles were inserted into a wide range of simulated orbits about Rhea in the restricted three-body problem framework (17). Over  $>35,000$  Saturnian years, semimajor axis, eccentricity, and inclination perturbations were remarkably restricted, as found in other contexts (18). Most  $<30^\circ$  inclination particles survived the full,  $10^5$ -year integration period if initial pericentres permitted small eccentricity excursions. Particles at the brief electron dropouts' distances were largely unperturbed. Overall stability was unchanged over a separate,  $10^5$ -year integration that additionally included Titan's presence and Saturn's oblateness (4).
16. R. J. Wagner *et al.*, *Lunar and Planetary Science Conference XXXVIII*, abstr. 1958, Lunar and Planetary Institute (League City, TX, 12 to 16 March 2007).
17. E. Everhart, in *Dynamics of Comets: Their Origin and Evolution*, A. Carusi, B. Valsecchi, Eds. (D. Reidel, Dordrecht, Netherlands, 1985), pp. 185–202.
18. D. P. Hamilton, J. A. Burns, *Icarus* **92**, 118 (1991).
19. S. M. Krimigis *et al.*, *Space Sci. Rev.* **114**, 233 (2004).
20. D. T. Young *et al.*, *Space Sci. Rev.* **114**, 1 (2004).
21. D. A. Gurnett *et al.*, *Space Sci. Rev.* **114**, 395 (2004).
22. M. K. Dougherty *et al.*, *Space Sci. Rev.* **114**, 331 (2004).
23. K. K. Khurana, C. T. Russell, M. K. Dougherty, *Icarus* **192**, 465 (2008).
24. W. S. Kurth *et al.*, *Planet. Space Sci.* **54**, 988 (2006).
25. E. Roussos *et al.*, *Icarus* **193**, 455 (2008).

26. Cassini-Huygens is a mission of international collaboration between NASA, the European Space Agency (ESA), and the Agenzia Spaziale Italiana (ASI). We thank M. Kusterer (The Johns Hopkins University Applied Physics Laboratory, JHUAPL), N. Achilleos, and S. Kellock (Imperial College London) for data reduction. MIMI/LEMMS was partly financed by the German Bundesministerium für Bildung und Forschung through the German Aerospace Center (DLR) and by the Max Planck

Gesellschaft. Work at JHUAPL, Johns Hopkins University, University of Iowa, and Southwest Research Institute was supported by NASA, as was that at Los Alamos National Laboratory, under the auspices of the U.S. DOE. Work at University College London and Imperial College London was supported by the U.K. Science and Technology Facilities Council. We acknowledge suggestions by T. V. Johnson and discussions with N. André and J. Schmidt.

#### Supporting Online Material

[www.sciencemag.org/cgi/content/full/319/5868/1384/DC1](http://www.sciencemag.org/cgi/content/full/319/5868/1384/DC1)

Materials and Methods

SOM Text

Figs. S1 to S9

References

9 October 2007; accepted 1 February 2008

10.1126/science.1151524

# TOPLESS Mediates Auxin-Dependent Transcriptional Repression During *Arabidopsis* Embryogenesis

Heidi Szemenyei,<sup>1,2</sup> Mike Hannon,<sup>1,2</sup> Jeff A. Long<sup>1\*</sup>

The transcriptional response to auxin is critical for root and vascular development during *Arabidopsis* embryogenesis. Auxin induces the degradation of AUXIN/INDOLE-3-ACETIC ACID (AUX/IAA) transcriptional repressors, freeing their binding partners, the AUXIN RESPONSE FACTOR (ARF) proteins, which can activate transcription of auxin response genes. We show that TOPLESS (TPL) can physically interact with IAA12/BODENLOS (IAA12/BDL) through an ETHYLENE RESPONSE FACTOR (ERF)-associated amphiphilic repression (EAR) motif. TPL can repress transcription *in vivo* and is required for IAA12/BDL repressive activity. In addition, *tpl-1* can suppress the patterning defects of the *bdl-1* mutant. Direct interaction between TPL and ARF5/MONOPTEROS, which is regulated by IAA12/BDL, results in a loss-of-function *arf5/mp* phenotype. These observations show that TPL is a transcriptional co-repressor and further our understanding of how auxin regulates transcription during plant development.

The phytohormone auxin has been implicated in many aspects of plant biology, including pattern formation during embryogenesis, lateral organ development, and cell expansion (1). Auxin mediates these activities through a signaling cascade involving AUXIN RESPONSE FACTORS (ARFs), which constitute a family of transcription factors that activate or repress expression of auxin response genes (2), and the AUX/IAA family of transcriptional repressors (3). AUX/IAAs negatively regulate auxin-mediated transcription by binding ARFs through conserved domains III and IV found in both types of proteins (4, 5).

Domain II of AUX/IAAs interacts with the auxin receptor TRANSPORT INHIBITOR RESISTANT (TIR1), part of a Skp1/cullin/F-box protein (SCF) complex (6–8). Auxin stabilizes this interaction, leading to the degradation of AUX/IAAs and allowing ARFs to activate transcription. Mutations in AUX/IAAs that disrupt the interaction with the TIR1 protein family abolish their auxin-induced degradation (8). For example, *bodenlos-1* (*bdl-1*) contains a stabilizing mutation (P74S, substitution of proline 74 with serine) in *IAA12* (9). IAA12/BDL physically in-

teracts with AUXIN RESPONSE FACTOR5/MONOPTEROS (ARF5/MP) and represses its activity (9). Consistent with this, both *bdl-1* mutants and loss-of-function *ARF5/MP* alleles display reduced vasculature and form a “basal peg” instead of a root and hypocotyl (the seedling stem) (10–12).

Although no prior connection has been made to auxin signaling, the most severe *topless-1* (*tpl-1*) phenotype, a homeotic transformation of the apical pole (the shoot) into a second basal pole (the root), suggests that it plays a role in this pathway (13). *tpl-1* is temperature sensitive and displays a range of phenotypes, all disrupting the patterning of the apical half of the embryo. The *tpl-1* mutation is a histidine substitution at asparagine 176 (N176H), and this allele acts as a dominant negative for the TOPLESS RELATED (TPR) family (14). On the basis of its domain structure and genetic interactions with a histone deacetylase and a histone acetyl transferase, TPL has been proposed to be a transcriptional co-repressor (14).

Transcriptional co-repressors do not bind DNA directly, but are recruited to DNA through interactions with DNA-binding transcription factors (15). Determining a co-repressor’s binding partners can provide insight into its biological relevance. Therefore, we performed yeast two-hybrid screens to find protein interactors using full-length TPL or the N terminus of TPL (Fig. 1A). Multiple AUX/IAAs, including IAA12/BDL, were among the positive clones in both screens (table S2). IAA12/BDL contains the four

conserved domains found in most AUX/IAAs (domains I to IV) (Fig. 1A). Through truncation analyses, we determined that domain I of IAA12/BDL is necessary and sufficient to interact with the C-terminal to lissencephaly homology (CTLH) domain of TPL (16) (Fig. 1, B and C).

Domain I of the AUX/IAAs contains an ERF-associated amphiphilic repression (EAR) motif (17), which is known to be involved in transcriptional repression (18). To determine if this motif is essential for the interaction with TPL, we tested an IAA12/BDL construct in which three leucines at the core of the EAR domain were changed to alanines (IAA12/BDL mEAR) (Fig. 1A). These changes to the EAR domain severely weakened the interaction between TPL and IAA12/BDL both in yeast two-hybrid assays (Fig. 1C) and assays in which glutathione S-transferase (GST)-tagged *in vitro*-translated products were used (Fig. 1D).

We also tested the interaction between IAA12/BDL and TPL using plant lysates. We found that column-bound GST-IAA12/BDL and GST-N-TPL could affinity purify TPL fused to a 6× hemagglutinin tag (TPL-HA) from plant extracts (Fig. 1E), whereas purified GST or beads alone could not, further indicating that TPL interacts with IAA12/BDL. These results also demonstrate that TPL can homodimerize, a characteristic shared with other transcriptional co-repressors (19). GST-N-TPL N176H could also pull down BDL P74S-green fluorescent protein (GFP) from plant lysates (Fig. 1F). Mutations in the EAR domain of GST-IAA12/BDL disrupted the interaction with TPL (Fig. 1E). These interactions were also observed in a tobacco transient expression system (20) by means of bimolecular fluorescence complementation (BiFC) (21) (fig. S2).

To investigate the biological significance of these physical interactions, we analyzed *tpl-1bdl-1* double mutants. *bdl-1* seedlings form a basal peg (Fig. 2C), lacking hypocotyl and root structures of wild-type seedlings (Fig. 2A), whereas *tpl-1* roots appear normal (Fig. 2B). *tpl-1bdl-1* seedlings formed hypocotyls and roots (Fig. 2D and table S3), indicating that *tpl-1* can suppress the basal patterning defects seen in *bdl-1*. *bdl-1* mutants also display a reduction in cotyledon (seed leaf) vasculature development (Fig. 2G). Wild-type and *tpl-1* cotyledons develop a primary midvein with loops of lateral veins along the margins (Fig. 2, E and F). In *bdl-1*, these veins are either absent or severely reduced ( $n = 70$ ) (Fig. 2G) (10). Vasculature formation was

<sup>1</sup>Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA. <sup>2</sup>Division of Biological Sciences, Section of Cell and Developmental Biology, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA.

\*To whom correspondence should be addressed. E-mail: [long@salk.edu](mailto:long@salk.edu)

restored in the cotyledons of *tpl-1bdl-1* seedlings, although lateral veins often did not form loops ( $n = 50$ ) (Fig. 2H). Therefore, *tpl-1* can suppress defects in both the apical and basal halves of *bdl-1* seedlings.

We then examined these genetic interactions during embryogenesis. In wild-type embryos, the hypophysis (the uppermost cell of the extra-embryonic suspensor) divides asymmetrically, resulting in a lens-shaped cell (Fig. 2I) that

will form the quiescent center of the root meristem. These cells express the synthetic auxin response reporter *DR5<sub>rev</sub>::GFP* (Fig. 2M) (4, 22), which is correlated with the accumulation of high auxin concentrations and is dependent on MP activity (22, 23). *tpl-1* embryos (Fig. 2J) also generate a lens-shaped cell and display a wild-type *DR5<sub>rev</sub>::GFP* expression pattern (Fig. 2N). *bdl-1* embryos, in which the hypophysis divides abnormally, lack the lens-shaped cell (Fig. 2K), and *DR5<sub>rev</sub>::GFP* expression is not detectable in hypophyseal cell derivatives ( $n = 50$ ) (Fig. 2O) (10, 23). In *tpl-1bdl-1* embryos, the lens-shaped cell is restored (Fig. 2L), and *DR5<sub>rev</sub>::GFP* expression is again detected in the developing embryonic root ( $n = 55$ ) (Fig. 2P). This rescue, combined with the interaction of TPL and the EAR domain of IAA12/BDL, suggests that TPL is involved in IAA12/BDL-mediated transcriptional repression. Therefore, we hypothesized that in the *tpl-1* background, IAA12/BDL cannot fully repress transcription of its target genes.

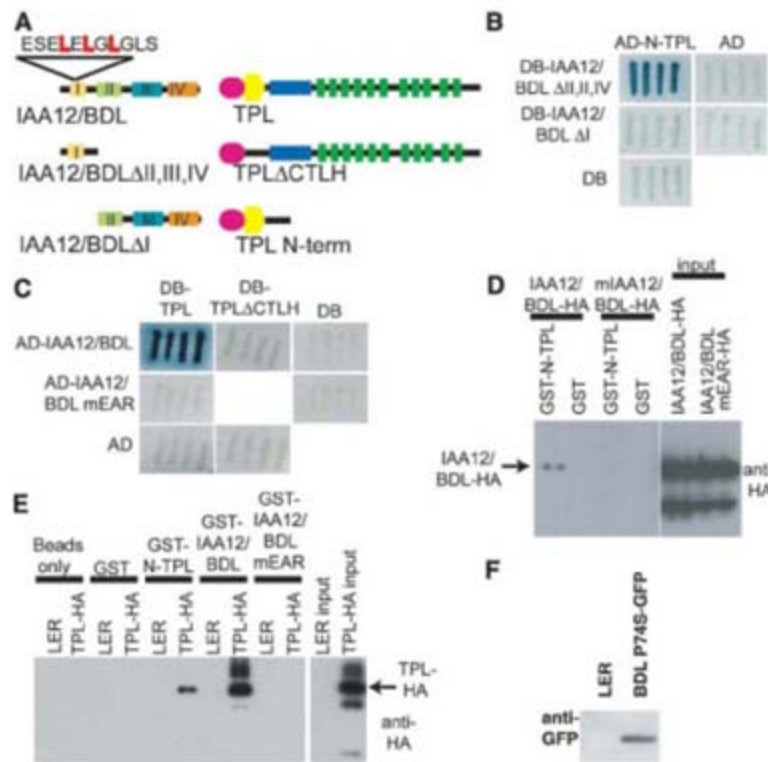
To test this hypothesis, we used a UAS/GAL4-based transcriptional repression assay in planta (Fig. 3A) (24). Plants expressing TPL-GAL4 and IAA12/BDL P74S-GAL4 fusions displayed a decrease in  $\beta$ -glucuronidase (GUS) activity in a wild-type background compared to the control line (Fig. 3B) ( $74.6 \pm 1.2\%$  and  $76.7 \pm 6.9\%$  of the control, respectively). Their representative staining patterns are shown in Fig. 3, C to F. The ability of IAA12/BDL P74S-GAL4 to repress transcription was diminished in the *tpl-1* background (Fig. 3, B and E). These data show that TPL can repress transcription and suggest the role of TPL as a transcriptional co-repressor in AUX/IAA-regulated transcriptional repression.

Single and multiple loss-of-function *aux/iaa* mutants are reported to have subtle or no discernible phenotypes (25). However, we hypothesized that further loss of AUX/IAA function might enhance the *tpl-1* phenotype. Therefore, we examined the effect of a transferred DNA (T-DNA) insertion allele (Salk\_138684) (26) of *IAA12/BDL* on *tpl-1*. This allele of *IAA12/BDL* has no obvious phenotype, whereas *tpl-1* exhibits a range of phenotypes when grown at 21°C (13). A higher frequency of severe seedling phenotypes was seen in the double-mutant background versus *tpl-1* alone when grown at 21°C (table S1), suggesting that decreased AUX/IAA function contributes to the *tpl-1* phenotype.

AUX/IAAs are proposed to bind ARFs and inhibit transcription of ARF target genes. Because IAA12/BDL binds ARF5/MP and also interacts with TPL, we reasoned that these three proteins act in a complex. We examined this possibility using BiFC (21) in a tobacco transient expression system (20). Although we observed no direct interaction between either TPL or TPL N176H and ARF5/MP (Fig. 4A and fig. S2I), interaction was observed upon coexpression of stabilized IAA12/BDL (Fig. 4, B and C). This suggests that these proteins can exist in a complex and that IAA12/BDL represses ARF5/MP by recruiting

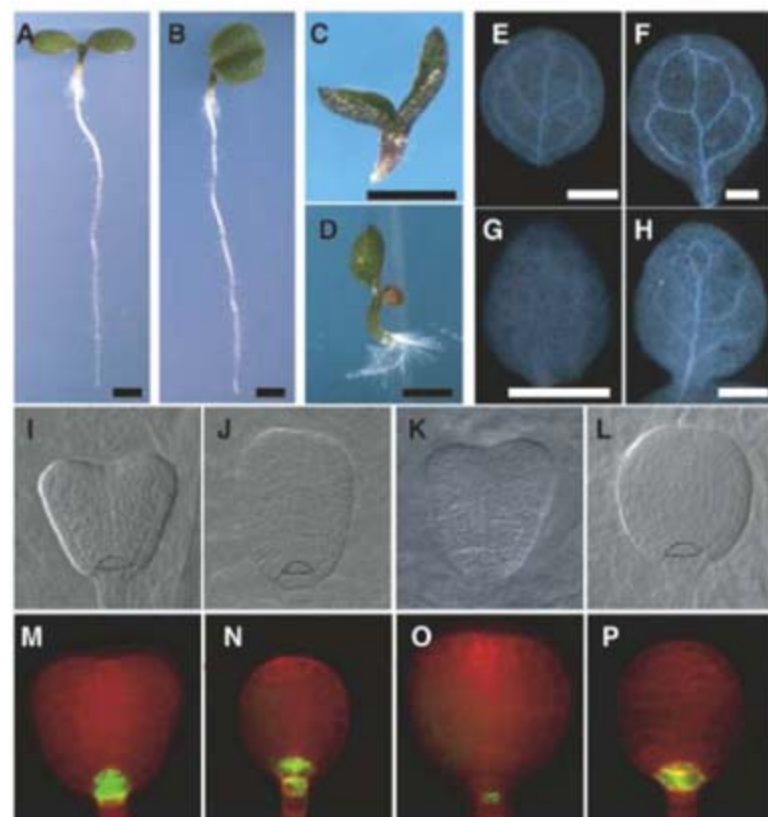
**Fig. 1.** Interactions between TPL and IAA12/BDL.

(A) Diagrams of TPL and IAA12/BDL constructs for interaction studies. Core residues of the EAR domain of IAA12/BDL are indicated in red. Pink ellipses represent LisH domains, yellow hexagons represent CTLH domains, blue boxes represent proline-rich regions, and green boxes represent WD40 repeats of TPL. (B) TPL interacts with domain I of IAA12/BDL in yeast. Interaction does not occur when domain I is truncated. (C) TPL and IAA12/BDL interact much less strongly when the EAR domain is mutated (IAA12/BDL mEAR) compared to the wild type. IAA12/BDL does not interact with TPL without the CTLH domain. AD, activation domain; DB, DNA binding domain. (D) The GST-N-TPL fusion protein can interact with in vitro-translated IAA12-HA, but does not interact strongly with IAA12/BDL mEAR. (E) Interaction results of extracts from nontransgenic or transgenic plants expressing TPL-HA with either beads alone or bound GST fusion proteins. LER indicates wild-type Landsberg *erecta* extracts. (F) GST-N-TPL N176H pulldown assay from nontransgenic or transgenic plants expressing BDL P74S-GFP. (See supplementary fig. 1, A and B, for loading controls.)



**Fig. 2.** *tpl-1* suppresses *bdl-1*.

(A to D) Seedlings, 4 days postgermination (dpg). Wild type (A), *tpl-1* (B), *bdl-1* (C), and *bdl-1tpl-1* (D). (E to H) Cleared cotyledons (4 dpg). Wild type (E), *tpl-1* (F), *bdl-1* (G), and *bdl-1tpl-1* (H). (I to L) Transition-stage embryos. Wild type (I), *tpl-1* (J), *bdl-1* (K), and *bdl-1tpl-1* (L). Lens-shaped cell and derivatives are outlined. (M to P) *DR5<sub>rev</sub>::GFP*-expressing embryos. Wild type (M), *tpl-1* (N), *bdl-1* (O), and *bdl-1tpl-1* (P). Scale bars: (A to D) 0.5 mm; (E to H) 1 mm.



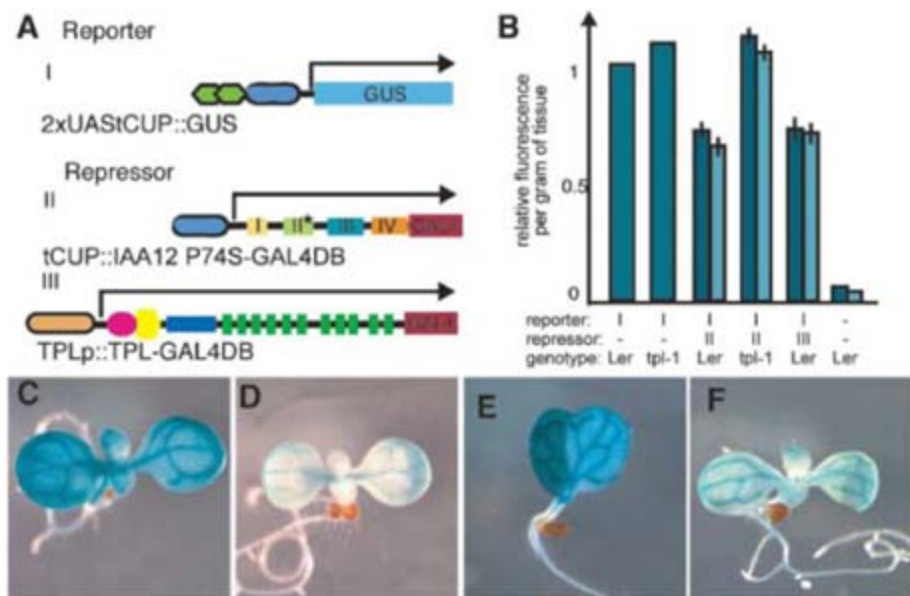
TPL. It also shows that TPL N176H does not disrupt this complex.

To determine if the EAR domain of IAA12/BDL allows IAA12/BDL to bridge TPL and ARF5/MP, we generated a chimeric protein containing domains III and IV of IAA12/BDL fused to the C terminus of TPL. This fusion protein should enable TPL to directly interact with ARF5/MP without requiring the EAR motif and should be unaffected by auxin due to the absence of domain II. First, we confirmed this interaction in a yeast two-hybrid assay (fig. S3), and then expressed this construct in plants under the control of the IAA12/BDL promoter. In 15

independent transgenic lines, we observed phenotypes similar to those of *bdl-1* and *arf5/mp* loss-of-function mutants (Fig. 4, D to F).

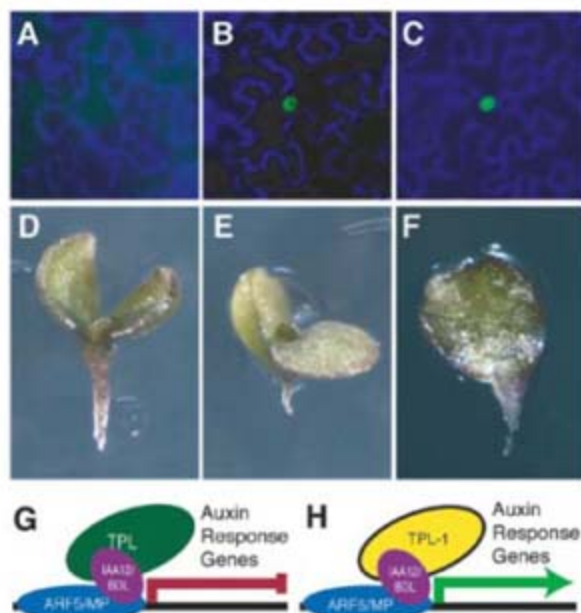
In light of these results, we propose a model in which TPL is a transcriptional co-repressor involved in the repression of auxin response genes through its physical interaction with the EAR domain of IAA12/BDL (Fig. 4G), and that this process is defective in *tpl-1*, resulting in the derepression of auxin response genes (Fig. 4H). Our results are also consistent with a report that application of a histone deacetylase inhibitor can suppress the lateral root defects seen in the stabilized *iaa14/solitary-root* mutant (27).

The coupling of TPL to ARFs through AUX/IAAs provides the plant an elegant mechanism to control ARF transcriptional activity in an auxin-dependent manner. It will be key to determine if all AUX/IAA proteins use TPL in this manner. It will also be of interest to determine if other EAR domain-containing transcription factors use TPL to facilitate their transcriptional repression activity in *Arabidopsis* as well as in other plant species.



**Fig. 3.** TPL can repress transcription and affects the ability of IAA12/BDL to repress. (A) Diagram of constructs analyzed in repressor assay. (B)  $\beta$ -Glucuronidase (GUS) activity measured by MUG assay (4-methylumbelliferyl glucuronide breakdown into 4-methylumbelliferone) of two independent transgenic repressor lines (dark blue/light blue). Experiment was done in duplicate. Repressor activity is relative to MUG activity of the reporter in wild-type Ler, where error bars show standard error of the mean. (C to F) GUS staining patterns in plants expressing reporter only (C), tCUPp::BDL P74S-GAL4DB and reporter in a wild-type background (D), tCUPp::BDL P74S-GAL4DB and reporter in the *tpl-1* background (E), and TPLp::TPL-GAL4DB and reporter in a wild-type Ler background (F).

**Fig. 4.** TPL functions to repress auxin response through its physical interaction with IAA12/BDL. (A to C) Bimolecular fluorescence complementation studies in tobacco. Tobacco transformed with TPL-SPYCE and ARF5/MP-SPYNE (A); BDL P74S, TPL-SPYCE, and ARF5/MP-SPYNE (B); or BDL P74S, TPL N176H-SPYCE, and ARF5/MP-SPYNE (C). (D to F) Phenotypes of transgenic plants expressing the TPL-IAA12/BDL III/IV fusion construct. Seedling lacking root (D); seedling lacking root and hypocotyl (E); seedling consisting of single cotyledon with no hypocotyl or root (F). (G) Model of TPL-mediated transcriptional repression of auxin response genes. (H) Schematic of derepression of auxin response genes in the *tpl-1* context.



References and Notes

- P. J. Davies, in *Plant Hormones: Biosynthesis, Signal Transduction, Action*, P. J. Davies, Ed. (Kluwer, Dordrecht, Netherlands, 2004), pp. 1–15.
- T. Ulmasov, G. Hagen, T. J. Guilfoyle, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5844 (1999).
- S. B. Tiwari, X. J. Wang, G. Hagen, T. J. Guilfoyle, *Plant Cell* **13**, 2809 (2001).
- T. Ulmasov, J. Murfett, G. Hagen, T. J. Guilfoyle, *Plant Cell* **9**, 1963 (1997).
- J. Kim, K. Harter, A. Theologis, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 11786 (1997).
- S. Kepinski, O. Leyser, *Nature* **435**, 446 (2005).
- N. Dharmasiri, S. Dharmasiri, M. Estelle, *Nature* **435**, 441 (2005).
- W. M. Gray, S. Kepinski, D. Rouse, O. Leyser, M. Estelle, *Nature* **414**, 271 (2001).
- T. Hamann, E. Benkova, I. Baurle, M. Kientz, G. Jurgens, *Genes Dev.* **16**, 1610 (2002).
- T. Hamann, U. Mayer, G. Jurgens, *Development* **126**, 1387 (1999).
- C. S. Hardtke, T. Berleth, *EMBO J.* **17**, 1405 (1998).
- T. Berleth, G. Jurgens, *Development* **118**, 575 (1993).
- J. A. Long, S. Woody, S. Poethig, E. M. Meyerowitz, M. K. Barton, *Development* **129**, 2797 (2002).
- J. A. Long, C. Ohno, Z. R. Smith, E. M. Meyerowitz, *Science* **312**, 1520 (2006).
- Z. Paroush et al., *Cell* **79**, 805 (1994).
- R. D. Emes, C. P. Ponting, *Hum. Mol. Genet.* **10**, 2813 (2001).
- S. B. Tiwari, G. Hagen, T. J. Guilfoyle, *Plant Cell* **16**, 533 (2004).
- M. Ohta, K. Matsui, K. Hiratsu, H. Shinshi, M. Ohme-Takagi, *Plant Cell* **13**, 1959 (2001).
- M. Pinto, C. G. Lobe, *J. Biol. Chem.* **271**, 33026 (1996).
- O. Voinnet, S. Rivas, P. Mestre, D. Baulcombe, *Plant J.* **33**, 949 (2003).
- M. Walter et al., *Plant J.* **40**, 428 (2004).
- J. Friml et al., *Nature* **426**, 147 (2003).
- D. Weijers et al., *Dev. Cell* **10**, 265 (2006).
- K. Wu, L. Tian, J. Hollingworth, D. C. Brown, B. Miki, *Plant Physiol.* **128**, 30 (2002).
- P. J. Overvoorde et al., *Plant Cell* **17**, 3282 (2005).
- J. M. Alonso et al., *Science* **301**, 653 (2003).
- H. Fukaki, N. Taniguchi, M. Tasaka, *Plant J.* **48**, 380 (2006).
- We thank D. Ernst for technical assistance, D. Weijers for *bdl-1* seeds, J. Friml for transgenic DRS<sub>uv</sub>::GFP lines, and the Arabidopsis Biological Resource Center for the SALK\_138684 insertion line. We also thank J. Chory, M. Yanofsky, J. Nemhauser, N. Krogan, and B. D. Gregory for critical examination of the manuscript. This work was supported by NIH grant GM072764 to J.A.L.

Supporting Online Material

[www.sciencemag.org/cgi/content/full/1151461/DC1](http://www.sciencemag.org/cgi/content/full/1151461/DC1)  
Materials and Methods  
Figs. S1 to S3  
Tables S1 to S3  
References

9 October 2007; accepted 21 January 2008  
Published online 7 February 2008;  
10.1126/science.1151461  
Include this information when citing this paper.

# De Novo Computational Design of Retro-Aldol Enzymes

Lin Jiang,<sup>1,2\*</sup> Eric A. Althoff,<sup>1\*</sup> Fernando R. Clemente,<sup>4</sup> Lindsey Doyle,<sup>5</sup> Daniela Röthlisberger,<sup>1</sup> Alexandre Zanghellini,<sup>1,2</sup> Jasmine L. Gallaher,<sup>3</sup> Jamie L. Betker,<sup>1</sup> Fujie Tanaka,<sup>6</sup> Carlos F. Barbas III,<sup>6</sup> Donald Hilvert,<sup>7</sup> Kendall N. Houk,<sup>4</sup> Barry L. Stoddard,<sup>5</sup> David Baker<sup>1,2,3,†</sup>

The creation of enzymes capable of catalyzing any desired chemical reaction is a grand challenge for computational protein design. Using new algorithms that rely on hashing techniques to construct active sites for multistep reactions, we designed retro-aldolases that use four different catalytic motifs to catalyze the breaking of a carbon-carbon bond in a nonnatural substrate. Of the 72 designs that were experimentally characterized, 32, spanning a range of protein folds, had detectable retro-aldolase activity. Designs that used an explicit water molecule to mediate proton shuffling were significantly more successful, with rate accelerations of up to four orders of magnitude and multiple turnovers, than those involving charged side-chain networks. The atomic accuracy of the design process was confirmed by the x-ray crystal structure of active designs embedded in two protein scaffolds, both of which were nearly superimposable on the design model.

Enzymes are excellent catalysts, and the ability to design new active enzymes could have applications in drug production (1), green chemistry (2), and bioremediation of xenobiotic pollutants (3). To date, most enzyme design efforts have used selection methodologies to retrieve very rare active catalysts from large libraries of candidate protein variants (4–7). Recent advances in computational protein design have made it possible to design new protein folds (8) and binding interactions (9) and have opened the door to the possibility of computationally designing enzymatic catalysts for any chemical reaction. Despite recent progress (10, 11), creating enzymes for chemical transformations not efficiently catalyzed by naturally occurring enzymes remains a major challenge. Here, we describe (i) general computational methods for constructing active sites for multistep reactions consisting of superimposed reaction intermediates and transition states (TS) surrounded by protein functional groups in orientations optimal for catalysis (Fig. 1) and (ii) the use of this methodology to design novel catalysts for a retro-aldol reaction in which a carbon-carbon bond is broken in a nonnatural (i.e., not found in biological systems) substrate: 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone (Fig. 2A) (12).

The first step in the computational design of an enzyme is to define one or more potential

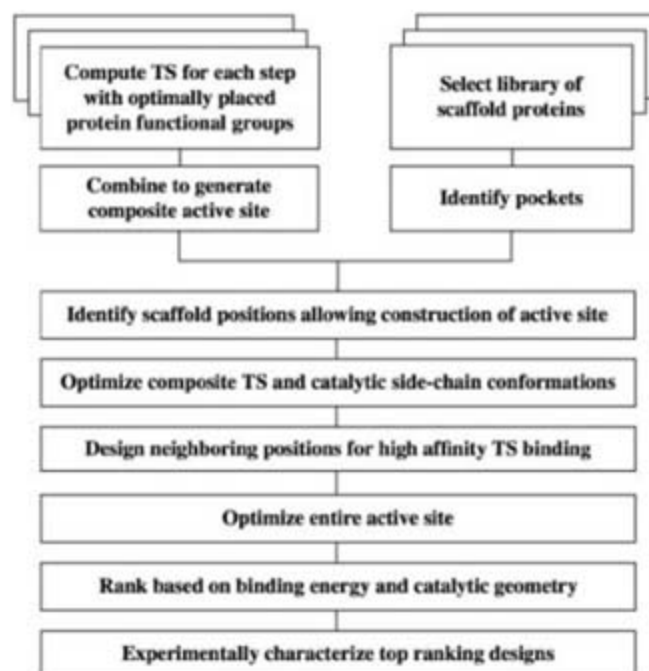
catalytic mechanisms for the desired reaction. For the retro-aldolase reaction, we focused on mechanisms involving enamine catalysis by lysine via a Schiff base or imine intermediate (13, 14). As shown in simplified form in Fig. 2B, the reaction proceeds in several distinct steps, involving acid-base catalysis by either amino acid side chains or water molecules. First, nucleophilic attack of lysine on the ketone of the sub-

strate forms a carbinolamine intermediate, which eliminates a water molecule to form the imine/iminium species. Next, carbon-carbon bond cleavage is triggered by the deprotonation of the  $\beta$ -alcohol, with the iminium acting as an electron sink. Finally, the enamine tautomerizes to an imine that is then hydrolyzed to release the covalently bound product and free the enzyme for another round of catalysis.

The second step of the design process is the identification of protein scaffolds that can accommodate the designed TS ensemble described above. To account for the multistep reaction pathway, we extended our enzyme design methodology (15) to allow the design of composite TS sites that are simultaneously compatible with multiple TS and reaction intermediates (16). Using this method, we generated design models using the four catalytic motifs shown schematically in Fig. 2C, which apply different constellations of catalytic residues to facilitate carbinolamine formation and water elimination, carbon-carbon bond cleavage, and release of bound product.

Because the probability of accurately reconstructing a given three-dimensional (3D) active site in an input protein scaffold is extremely small, it is essential to consider a very large set of active-

**Fig. 1.** Computational enzyme design protocol for a multistep reaction. The first step is to generate ensembles of models of each of the key intermediates and transition states in the reaction pathway in the context of a specific catalytic motif composed of protein functional groups. These models are then superimposed, based on the protein functional group positions, to create an initial composite active-site description. Large ensembles of distinct 3D realization of these composite active sites are then generated by simultaneously varying the degrees of freedom of the composite TS, the orientation of the catalytic side chains relative to the composite TS, and the internal conformation of the catalytic side chains. For each composite active-site description, candidate catalytic sites are gener-



ated in an input scaffold set by RosettaMatch (15). Briefly, each rotamer of each catalytic side chain is placed at each position in each scaffold, and the ensuing position of the composite TS is recorded in the hash. After the filling out of the hash table, which is linear in the numbers of scaffold positions and catalytic rotamers, the hash is searched for TS positions that are compatible with all catalytic constraints; such positions are termed "matches." For each match, the rigid body orientation of the composite TS and the internal coordinates of the catalytic side chains are optimized to reduce steric clashes while maintaining the catalytic geometry within specified tolerances. The remaining positions (not included in the minimal catalytic site description) surrounding the docked composite TS model are redesigned to optimize TS binding affinity by means of the standard Rosetta design methodology (20, 21). The rigid body orientation of the composite TS, the side-chain torsion angles, and (in some cases) the backbone torsion angles in the active site are refined via quasi-Newton optimization (22). The resulting designs are ranked based on the total binding energy to the composite TS and the satisfaction of the specified catalytic geometry, and then the top-ranked designs are experimentally characterized. The SOM contains detailed descriptions of each step in the protocol.

<sup>1</sup>Department of Biochemistry, University of Washington, Seattle, WA 98195, USA. <sup>2</sup>Biomolecular Structure and Design Program, University of Washington, Seattle, WA 98195, USA. <sup>3</sup>Howard Hughes Medical Institute (HHMI), University of Washington, Seattle, WA 98195, USA. <sup>4</sup>Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA. <sup>5</sup>Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. <sup>6</sup>The Skaggs Institute for Chemical Biology and the Departments of Chemistry and Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA. <sup>7</sup>Laboratory of Organic Chemistry, Eidgenössische Technische Hochschule (ETH) Zürich, 8093 Zürich, Switzerland.

\*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: dbaker@u.washington.edu



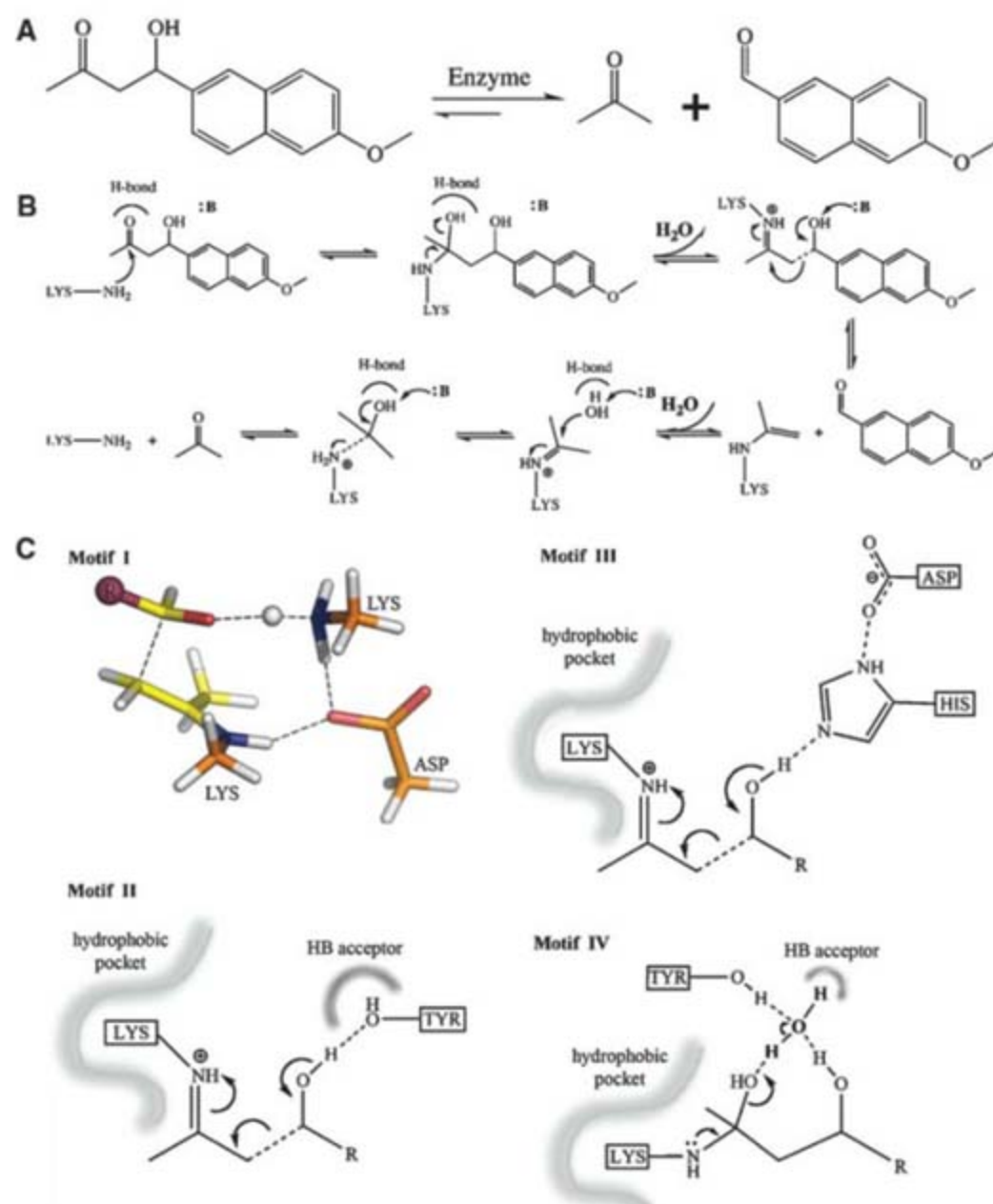
site possibilities. We generated such a set by simultaneously varying (i) the internal degrees of freedom of the composite TS (fig. S1B), (ii) the orientation of the catalytic side chains with respect to the composite TS (fig. S3), within ranges that are consistent with catalysis, and

(iii) the conformations of the catalytic side chains (fig. S3). For example, in a representative calculation for motif III, we searched for placements of a total  $1.4 \times 10^{18}$  possible 3D active sites (table S3) at all triples or quadruples of backbone positions surrounding binding pockets in 71 different

protein scaffolds (table S4). This combinatorial matching resulted in a total of 181,555 distinct solutions for the placement of the composite TS and the surrounding catalytic residues. Through extensive pruning at multiple levels, and by breaking the combinatoric explosion via hashing, the RosettaMatch algorithm (15) is able to rapidly eliminate most active-site possibilities in a given scaffold that are unfavorable as a result of poor catalytic geometry or significant steric clashes with very little computational cost. After optimization of the composite TS rigid body orientation and the identities and conformations of the surrounding residues, a total of 72 designs with 8 to 20 amino acid identity changes in 10 different scaffolds were selected for experimental characterization based on the predicted TS binding energy, the extent of satisfaction of the catalytic geometry, the packing around the active lysine, and the consistency of side-chain conformation after side-chain repacking in the presence and absence of the TS model (16). Genes encoding the designs were synthesized and the proteins were expressed and purified from *Escherichia coli*; soluble purified protein was obtained for 70 of the 72 expressed designs.

Retro-aldolase activity was monitored via a fluorescence-based assay of product formation (12) for each of the designs, and the results are summarized in Table 1. Our initial 12 designs used the first active site shown in Fig. 2C, which involves a charged side-chain (Lys-Asp-Lys)-mediated proton transfer scheme resembling that in D-2-deoxyribose-5-phosphate aldolase (13). Of these designs, two showed slow enaminone formation with 2,4-pentandione (17), which is indicative of a nucleophilic lysine, but none displayed retro-aldolase activity (16). Ten designs were made for the second, much simpler active site shown in Fig. 2C, which involves a single imine-forming lysine in a hydrophobic pocket similar to aldolase catalytic antibodies; of these designs, one formed the enaminone, but none were catalytically active. The third active site incorporates a His-Asp dyad as a general base to abstract a proton from the  $\beta$ -alcohol; of the 14 designs tested, 10 exhibited stable enaminone formation, and 8 had detectable retro-aldolase activity. In the final active site, we experimented with the explicit modeling of a water molecule, positioned via side-chain hydrogen-bonding groups, which shuttles between stabilizing the carbinolamine and abstracting the proton from the hydroxyl. Of the 36 designs tested, 20 formed the enaminone and 23 (with 11 distinct positions for the catalytic lysine) had significant retro-aldolase activity, with rate enhancements up to four orders of magnitude over the uncatalyzed reaction (18).

The active designs occur on five different protein scaffolds belonging to the triose phosphate isomerase (TIM)-barrel and jelly-roll folds. The most active designs exhibited multiple turn-over kinetics; the linear progress curves for de-



**Fig. 2.** Retro-aldol reaction and active-site motifs. (A) The retro-aldol reaction. (B) General description of the aldol reaction pathway with a nucleophilic lysine and general acid-base chemistry. Several of the proton transfer steps are left out for brevity. (C) Active-site motifs with quantum mechanically optimized structures (23). (Top left) Motif I. Two lysines are positioned nearby one another to lower the  $pK_a$  of the nucleophilic lysine, and a Lys-Asp dyad acts as the base to deprotonate the hydroxyl group. (Bottom left) Motif II. The catalytic lysine is buried in a hydrophobic environment to lower its  $pK_a$  to make it a more potent nucleophile, and a tyrosine functions as a general acid or base. HB, hydrogen-bond. (Top right) Motif III. The catalytic lysine, analogous to motif II, is placed in a hydrophobic pocket to alter its  $pK_a$ , and a His-Asp dyad serves as a general base similar to the catalytic unit commonly observed in the serine proteases (24). (Bottom right) Motif IV. The catalytic lysine is again positioned in a hydrophobic environment. Additionally, an explicitly modeled bound water molecule is placed such that it forms a hydrogen bond with the carbinolamine hydroxyl during its formation, aids in the water elimination step, and deprotonates the  $\beta$ -alcohol at the carbon-carbon bond-breaking step. A hydrogen-bond donor/acceptor, such as Ser, Thr, or Tyr, is placed to position the water molecule in a tetrahedral geometry with the  $\beta$ -alcohol and the carbinolamine hydroxyl. The proton-abstracting ability of the water molecule is enhanced by a second hydrogen bond with a base residue. We incorporated, where possible, additional hydrogen-bonding interactions to stabilize the carbinolamine hydroxyl group and an aromatic side chain to optimally pack along the planar aromatic moiety of the substrate.

signs RA60 and RA61, for example, continue unchanged for more than 20 turnovers. Progress curves [Fig. 3A and supporting online material (SOM)] show a range of kinetic behaviors: In some cases (RA45), there is a pronounced lag phase, likely associated with slow imine formation, whereas in others (RA61), there is little or no lag, and for a third set, there is an initial burst followed by a slower steady-state rate (RA22). Notably, simple linear kinetics are observed for the designs in the relatively open jelly-roll scaffold, whereas more complex kinetics are observed for the TIM-barrel designs, which have more enclosed active-site pockets that may restrict substrate access and product release. To obtain  $k_{\text{cat}}$  and  $K_M$  estimates for several of the best enzymes (Fig. 3B), we extracted reaction velocities from the steady-state portions of the

progress curves and assumed simple Michaelis-Menten kinetics. Given the simplifications, these values are best viewed as phenomenological; future characterization will be required to define rate constants in a particular kinetic model. The apparent  $k_{\text{cat}}$  and  $K_M$  values are given in Table 2;  $k_{\text{uncat}}$  was determined from measurements of the reaction progression in the absence of enzyme and is close to previously determined values (18).  $k_{\text{cat}}/k_{\text{uncat}}$  for the most active designs is  $2 \times 10^4$ . The catalytic proficiency of the designs is far from that of naturally occurring enzymes, which have a  $k_{\text{cat}}/K_M$  of about  $1 \text{ M}^{-1} \text{ s}^{-1}$  (Table 2); the very low  $k_{\text{cat}}$  value is probably associated with low reactivity of the imine-forming lysine. Rates for all active designs with 270  $\mu\text{M}$  substrate are reported in table S1. For each of the 11 catalytic lysine positions, a “knockout” mu-

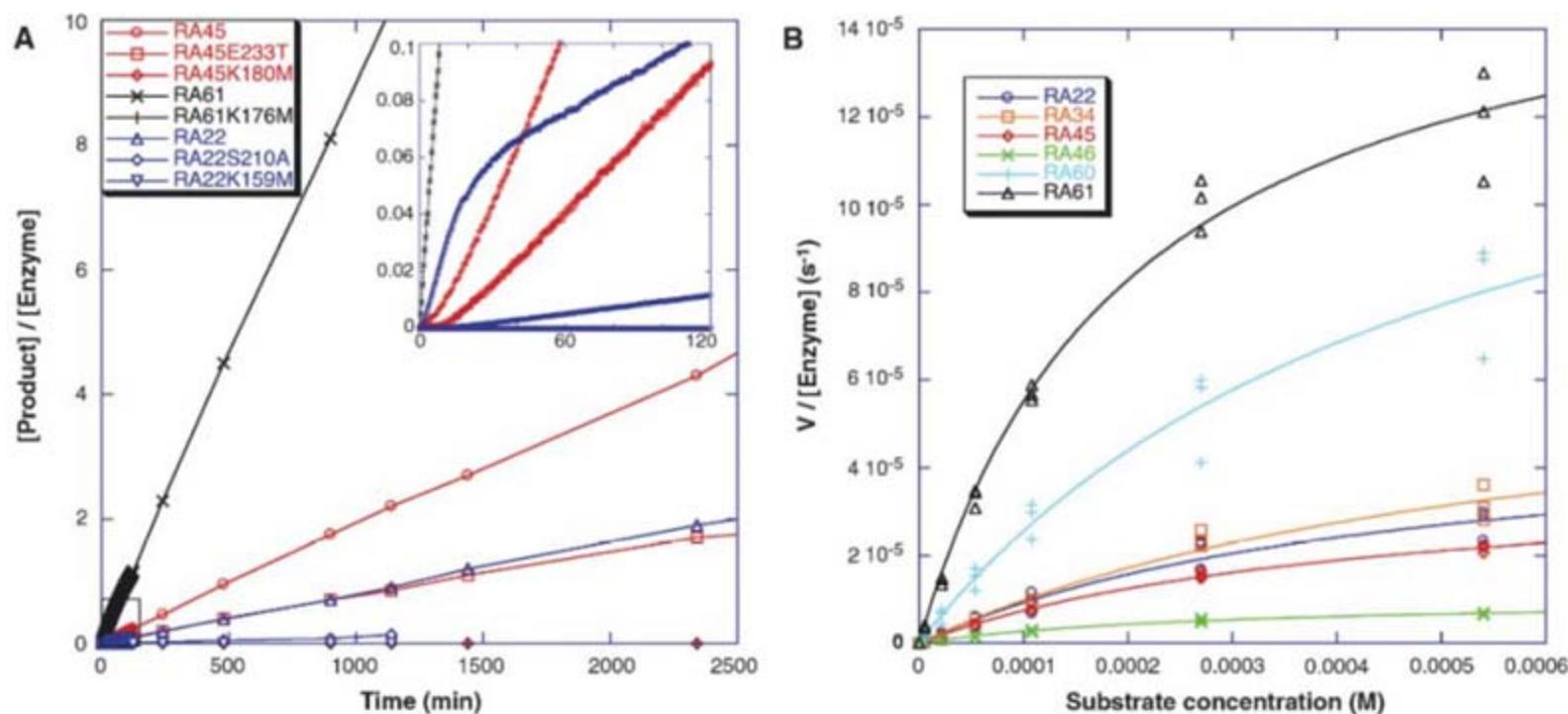
tation to methionine dramatically decreased the activity or, more commonly, abolished catalysis completely, verifying that the observed activity was due to the designed active site.

Design models for several of the most active designs with catalytic motif IV are shown in Fig. 4, A to C. Design RA60 (Fig. 4A) is on a jelly-roll scaffold, and RA45 (Fig. 4C) and RA46 (Fig. 4B) are on a TIM-barrel scaffold. The imine-forming lysine, the hydrogen-bonding residues coordinating the bridging water molecules, and the designed hydrophobic pocket (which binds the aromatic portion of the substrate) are clearly evident in all three designs.

To evaluate the accuracy of the design models, we solved the structures of two of the designs by x-ray crystallography (Fig. 4, D and E). The 2.2 Å resolution structure of the Ser<sup>210</sup>→Ala<sup>210</sup> (S210A) variant of RA22 (Fig. 4D) (19) shows that the designed catalytic residues Lys<sup>159</sup>, His<sup>233</sup>, and Asp<sup>53</sup> superimpose well on the original design model, and the remainder of the active site is nearly identical to the design. The 1.9 Å resolution structure of the M48K variant of RA61 likewise reveals an active site very close to that of the design model, with only His<sup>46</sup> and Trp<sup>178</sup> in alternative rotamer conformations, perhaps resulting from the absence of substrate in the crystal structure (Fig. 4E). Both crystal structures differ most significantly from the designs in the loops surrounding the active site; explicitly incorporating backbone flexibility in these regions

**Table 1.** Enaminone formation and enzyme activity for different active-site motifs. NC, not considered.

Motif	Catalytic lysine environment	Carbinolamine stabilization	Proton abstraction	Number tested	Number forming enaminone	Number of active designs	Rate enhancement
I	Polar	NC	Lys-Asp dyad	12	2	0	<4
II	Hydrophobic	NC	Tyr	9	1	0	<4
III	Hydrophobic	H-bond acceptor/donor	His-Asp dyad	13	10	10	$10^2$ – $10^3$
IV	Hydrophobic	Water, H-bond acceptor	Water	38	20	22	$10^3$ – $10^4$



**Fig. 3.** Experimental characterization of active enzyme designs. **(A)** Progress curves for RA61, RA61 K176M, RA22, RA22 S210A, RA22 K159M, RA45, RA45 E233T, and RA45 K180M. The enzymes were tested with 540  $\mu\text{M}$  of the racemic substrate; the reaction was followed by measuring the appearance of the fluorescent product (excitation wavelength, 330 nm; emission wavelength, 452 nm). The y axis is the concentration of product (determined from the fluorescence signal by a standard curve prepared with pure product solutions) divided by the enzyme concentration. In the design models, the serine-to-

alanine mutation in RA22 and the glutamate-to-threonine mutation in RA45 eliminate interactions that stabilize the carbinolamine intermediate and position the bound water molecule; both mutations reduce the reaction rate considerably. Mutation of the catalytic lysine residues to methionine completely eliminates enzyme activity. **(B)** Dependence of reaction velocity ( $V$ ) on substrate concentration. The rates are reported in Table 2. Reaction conditions for all experiments were 25 mM HEPES, 2.7%  $\text{CH}_3\text{CN}$ , 100 mM NaCl (pH 7.5), and substrate at the indicated concentration.

during the design process could yield improved enzymes in the future.

Each proposed catalytic mechanism can be treated as an experimentally testable hypothesis

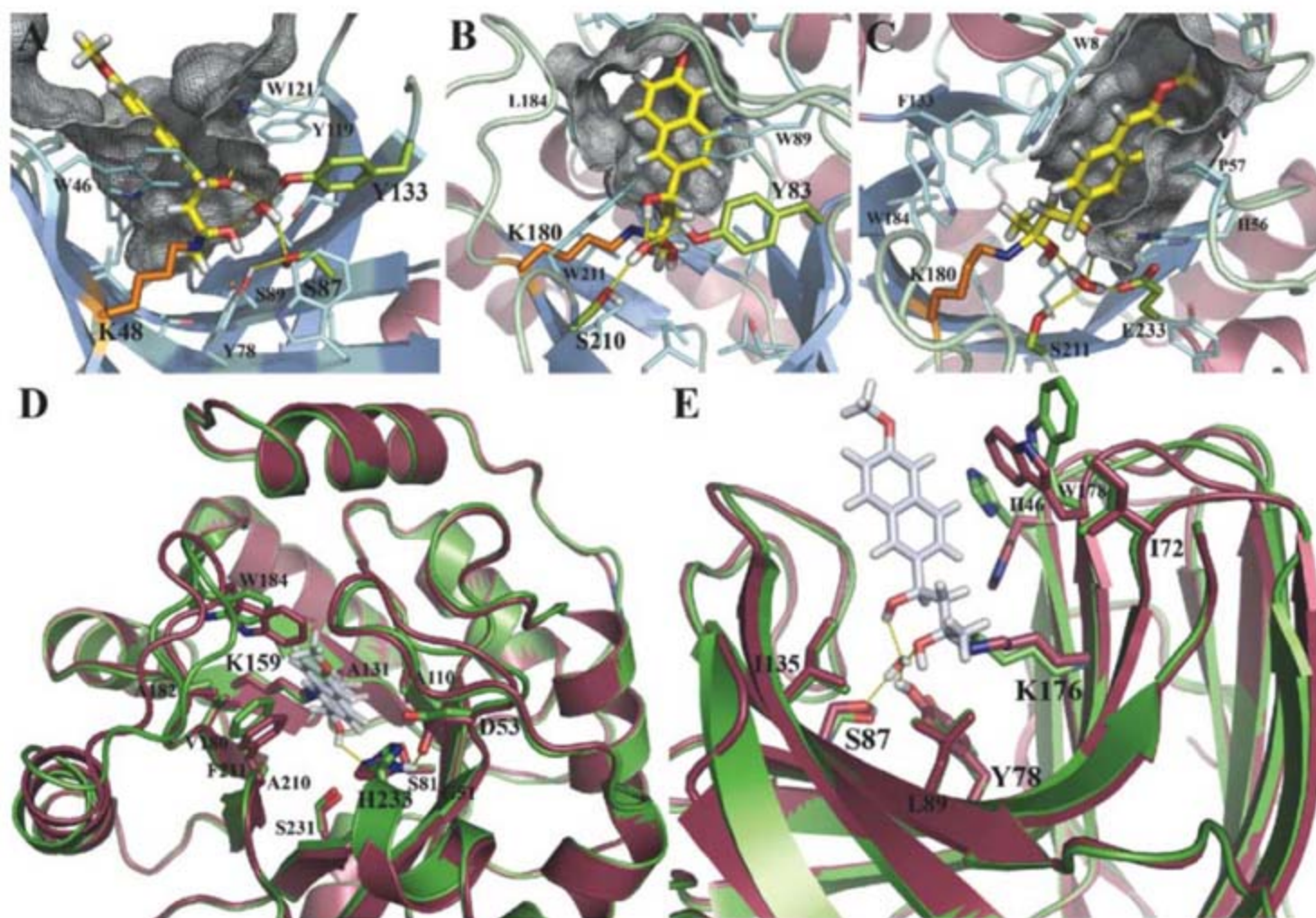
to be tested by multiple independent design experiments. Our lack of success with the first active sites that were tested contrasts markedly with our relatively high success rate with the active

site in which proton shuffling is carried out by a bound water molecule rather than by amino acid side chains acting as acid-base catalysts. The charged polar networks in highly optimized naturally occurring enzymes require exquisite control over functional group positioning and protonation states, as well as the satisfaction of the hydrogen-bonding potential of the buried polar residues, which leads to still more extended hydrogen-bond networks. Computational design of such extended polar networks is exceptionally challenging because of the difficulty of accurately computing the free energies of buried polar interactions, particularly the influence of polarizability on electrostatic free energies and the delicate balance between the cost of desolvation and the gain in favorable intraprotein electrostatic and hydrogen-bonding interactions. The sampling problem also becomes increasingly formidable

**Table 2.** Kinetic parameters of selected designs. b, burst phase; s, steady state.

Design	$k_{cat}$ ( $\times 10^{-3} \text{ min}^{-1}$ )	$K_M$ ( $\mu\text{M}$ )	$k_{cat}/K_M$ ( $\text{M}^{-1} \text{ s}^{-1}$ )	$k_{cat}/k_{uncat}^*$
RA22	$3.1 \pm 0.3$ (b) $0.5 \pm 0.1$ (s)	$480 \pm 130$ (b) $450 \pm 210$ (s)	$0.11 \pm 0.03$ (b) $0.018 \pm 0.006$ (s)	$8.1 \times 10^3$ (b) $1.2 \times 10^3$ (s)
RA34	$4.2 \pm 1.1$ (b) $0.6 \pm 0.1$ (s)	$620 \pm 180$ (b) $600 \pm 140$ (s)	$0.11 \pm 0.01$ (b) $0.016 \pm 0.004$ (s)	$1.1 \times 10^4$ (b) $1.5 \times 10^3$ (s)
RA45	$2.3 \pm 0.2$	$430 \pm 48$	$0.091 \pm 0.004$	$6.0 \times 10^3$
RA46	$0.62 \pm 0.5$	$290 \pm 60$	$0.037 \pm 0.006$	$1.6 \times 10^3$
RA60	$9.3 \pm 0.9$	$510 \pm 33$	$0.30 \pm 0.06$	$2.4 \times 10^4$
RA61	$9.0 \pm 1.0$	$210 \pm 50$	$0.74 \pm 0.11$	$2.3 \times 10^4$

\* $k_{uncat} = 3.9 \times 10^{-7} \text{ min}^{-1}$  (25).



**Fig. 4.** Structures of designed enzymes. (A to C) Examples of design models for active designs highlighting groups important for catalysis. The nucleophilic imine-forming lysine is in orange, the TS model is in yellow, the hydrogen-bonding groups are in light green, and the catalytic water is shown explicitly. The designed hydrophobic binding site for the aromatic portion of the TS model is indicated by the gray mesh. (A) RA60 (catalytic motif IV, jelly-roll scaffold). A designed hydrophobic pocket encloses the aromatic portion of the substrate and packs the aliphatic portion of the imine-forming Lys<sup>48</sup>. A designed hydrogen-bonding network positions the bridging water molecule and the composite TS. (B) RA46 (catalytic motif IV, TIM-barrel scaffold). Tyr<sup>83</sup> and Ser<sup>210</sup> position the bridging water molecule, which facilitates the proton shuffling required in active site IV. (C) RA45 (catalytic motif IV, TIM-barrel scaffold). The bridging water is hydrogen-bonded by Ser<sup>211</sup> and Glu<sup>233</sup>,

replacing the Glu<sup>233</sup> with Thr decreases catalytic activity threefold (Fig. 3A). (D and E) Overlay of design model (purple) on x-ray crystal structure (green). Designed amino acid side chains are shown in stick representation, and the TS model in the design is shown in gray. (D) The 2.2 Å crystal structure of the S210A variant of RA22 (catalytic motif III, TIM-barrel scaffold). The  $C_\alpha$  root mean square deviation (RMSD) between the design model and crystal structure is 0.62 Å, and the heavy-atom RMSD in the active site is 1.10 Å. (E) 1.8 Å crystal structure of M48K variant of RA61 (catalytic motif IV, jelly-roll scaffold). Design-crystal structure  $C_\alpha$  RMSD is 0.46 Å, and heavy-atom RMSD is 0.8 Å. The small differences in the high-resolution details of packing around the binding pocket and two rotamer changes in RA61 that may reflect the absence of a TS analog in the crystal structure.

for more complex sites: The side-chain identity and conformation combinatorics dealt with by hashing in RosettaMatch become intractable for sites consisting of five or more long polar side chains, which for accurate representation may require as many as 1000 rotamer conformations each. At the other extreme, bound water molecules offer considerable versatility, because they can readily reorient to switch between acting as hydrogen-bond acceptors and donors and involve neither delicate free-energy tradeoffs nor intricate interaction networks.

It is tempting to speculate that our computationally designed enzymes resemble primordial enzymes more than they resemble highly refined modern-day enzymes. The ability to design simultaneously only three to four catalytic residues parallels the infinitesimal probability that, early in evolution, more than three to four residues would have happened to be positioned appropriately for catalysis; some of the functions played by exquisitely positioned side chains in modern enzymes may have been played by water molecules earlier in evolutionary history.

Although our results demonstrate that novel enzyme activities can be designed from scratch and indicate the catalytic strategies that are most accessible to nascent enzymes, there is still a significant gap between the activities of our designed catalysts and those of naturally occurring enzymes. Narrowing this gap presents an exciting prospect for future work: What additional features have to be incorporated into the design process to achieve catalytic activities approaching those of naturally occurring enzymes? The close agreement between the two crystal structures and the design models gives credence to our strategy of testing

hypotheses about catalytic mechanisms by generating and testing the corresponding designs; indeed, almost any idea about catalysis can be readily tested by incorporation into the computational design procedure. Determining what is missing from the current generation of designs and how it can be incorporated into a next generation of more active designed catalysts will be an exciting challenge that should unite the fields of enzymology and computational protein design in the years to come.

#### References and Notes

1. D. K. Ro *et al.*, *Nature* **440**, 940 (2006).
2. O. Kirk, T. V. Borchert, C. C. Fuglsang, *Curr. Opin. Biotechnol.* **13**, 345 (2002).
3. D. B. Janssen, I. J. Dinkla, G. J. Poelarends, P. Terpstra, *Environ. Microbiol.* **7**, 1868 (2005).
4. D. Hilvert, *Annu. Rev. Biochem.* **69**, 751 (2000).
5. B. Seelig, J. W. Szostak, *Nature* **448**, 828 (2007).
6. F. H. Arnold, A. A. Volkov, *Curr. Opin. Chem. Biol.* **3**, 54 (1999).
7. O. Khersonsky, C. Roodveldt, D. S. Tawfik, *Curr. Opin. Chem. Biol.* **10**, 498 (2006).
8. B. Kuhlman *et al.*, *Science* **302**, 1364 (2003).
9. L. L. Looger, M. A. Dwyer, J. J. Smith, H. W. Hellinga, *Nature* **423**, 185 (2003).
10. D. N. Bolon, S. L. Mayo, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14274 (2001).
11. J. Kaplan, W. F. DeGrado, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11566 (2004).
12. F. Tanaka, R. Fuller, H. Shim, R. A. Lerner, C. F. Barbas III, *J. Mol. Biol.* **335**, 1007 (2004).
13. A. Heine *et al.*, *Science* **294**, 369 (2001).
14. S. W. Fullerton *et al.*, *Bioorg. Med. Chem.* **14**, 3002 (2006).
15. A. Zanghellini *et al.*, *Protein Sci.* **15**, 2785 (2006).
16. Materials and methods are available as supporting material on Science Online.
17. J. Wagner, R. A. Lerner, C. F. Barbas III, *Science* **270**, 1797 (1995).
18. F. Tanaka, C. F. Barbas III, *J. Am. Chem. Soc.* **124**, 3510 (2002).
19. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

20. G. Dantas, B. Kuhlman, D. Callender, M. Wong, D. Baker, *J. Mol. Biol.* **332**, 449 (2003).
21. J. Meiler, D. Baker, *Proteins* **65**, 538 (2006).
22. W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes in FORTRAN: The Art of Scientific Computing* (Cambridge Univ. Press, Cambridge, ed. 2, 1992).
23. F. R. Clemente, K. N. Houk, *J. Am. Chem. Soc.* **127**, 11294 (2005).
24. C. T. Porter, G. J. Bartlett, J. M. Thornton, *Nucleic Acids Res.* **32**, D129 (2004).
25. G. Zhong *et al.*, *Angew. Chem. Int. Ed. Engl.* **37**, 2481 (1998).
26. Kinetic parameters of the designs reported here were determined at the University of Washington. For selected designs, the kinetic parameters were confirmed by independent experiments performed at the Scripps Research Institute. We thank R. Fuller for technical assistance. Thorough testing of the four catalytic motifs was made possible through gene synthesis by Codon Devices. We thank Rosetta@Home participants for their valuable contributions of computer time. E.A.A. is funded by a Ruth L. Kirschstein National Research Service Award. This work was supported by the Defense Advanced Research Projects Agency and HHMI. Coordinates and structure factors for the crystal structures of RA22 variant S210A and RA61 variant M48K were deposited with the Research Collaboratory for Structural Bioinformatics Protein Data Bank (PDB) under the accession numbers 3B5V and 3B5L, respectively. The xyz coordinates of the designs RA22, RA34, RA45, RA46, RA60, and RA61 are included with the SOM as a zipped archive.

#### Supporting Online Material

[www.sciencemag.org/cgi/content/full/319/5868/1387/DC1](http://www.sciencemag.org/cgi/content/full/319/5868/1387/DC1)

Materials and Methods

SOM Text

Figs. S1 to S8

Tables S1 to S8

References

Design Model Coordinates in PDB Format

7 November 2007; accepted 5 February 2008

10.1126/science.1152692

## A Cholesterol Biosynthesis Inhibitor Blocks *Staphylococcus aureus* Virulence

Chia-I Liu,<sup>1,2,3\*</sup> George Y. Liu,<sup>4\*</sup> Yongcheng Song,<sup>5\*</sup> Fenglin Yin,<sup>6</sup> Mary E. Hensler,<sup>7</sup> Wen-Yih Jeng,<sup>1,2</sup> Victor Nizet,<sup>7,8†</sup> Andrew H.-J. Wang,<sup>1,2,3†</sup> Eric Oldfield<sup>5,6†</sup>

*Staphylococcus aureus* produces hospital- and community-acquired infections, with methicillin-resistant *S. aureus* posing a serious public health threat. The golden carotenoid pigment of *S. aureus*, staphyloxanthin, promotes resistance to reactive oxygen species and host neutrophil-based killing, and early enzymatic steps in staphyloxanthin production resemble those for cholesterol biosynthesis. We determined the crystal structures of *S. aureus* dehydrosqualene synthase (CrtM) at 1.58 angstrom resolution, finding structural similarity to human squalene synthase (SQS). We screened nine SQS inhibitors and determined the structures of three, bound to CrtM. One, previously tested for cholesterol-lowering activity in humans, blocked staphyloxanthin biosynthesis in vitro (median inhibitory concentration ~100 nM), resulting in colorless bacteria with increased susceptibility to killing by human blood and to innate immune clearance in a mouse infection model. This finding represents proof of principle for a virulence factor–based therapy against *S. aureus*.

Over the past 20 years, there has been an explosion in the prevalence of antibiotic-resistant bacterial infections, both in the

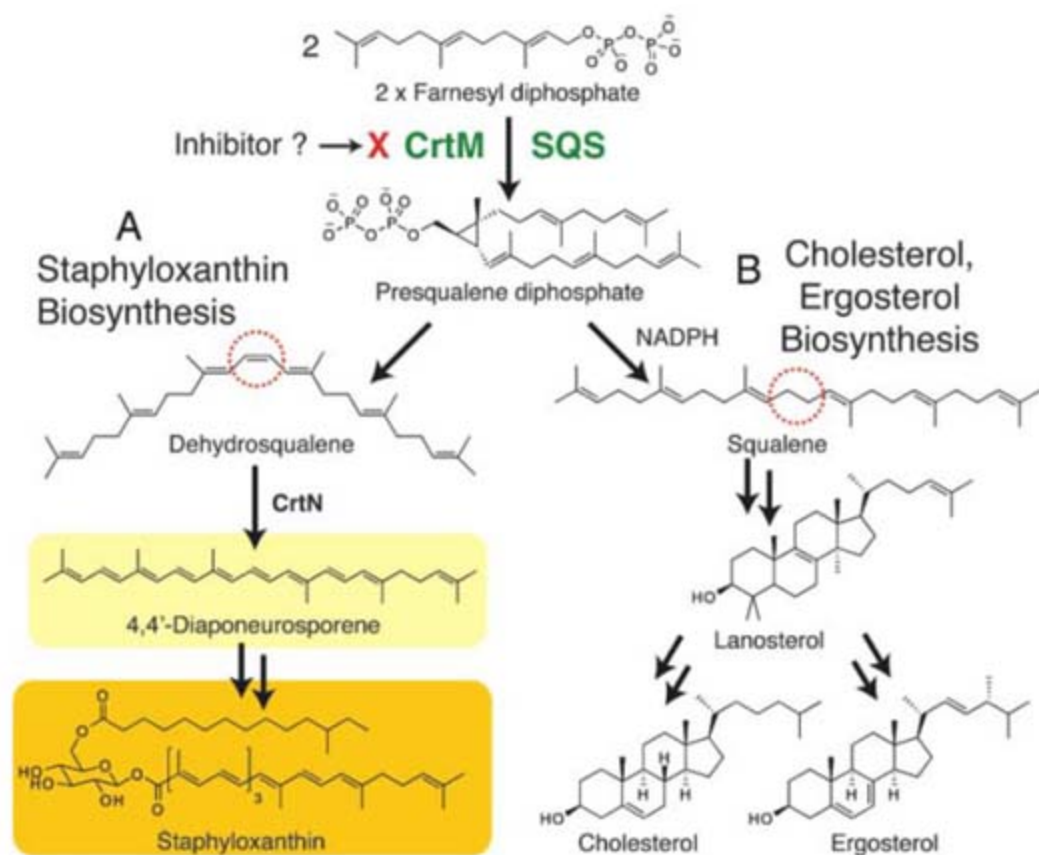
hospital and in the general community; in the United States, more deaths are now attributable to methicillin-resistant *Staphylococcus aureus*

(MRSA) infections than to HIV/AIDS (1, 2). Unfortunately, over the same time, there has been a decrease in the rate of discovery of new antibiotics, creating a pressing need for the development of novel infectious disease therapies. One approach, discussed at a recent National Research Council workshop (3), involves the specific neutralization of bacterial virulence factors to render pathogenic bacteria susceptible to innate immune system clearance. However,

<sup>1</sup>Institute of Biological Chemistry, Academia Sinica, Nankang, Taipei 11529, Taiwan. <sup>2</sup>National Core Facility of High-Throughput Protein Crystallography, Academia Sinica, Nankang, Taipei 11529, Taiwan. <sup>3</sup>Institute of Biochemical Sciences, College of Life Science, National Taiwan University, Taipei 10098, Taiwan. <sup>4</sup>Division of Pediatric Infectious Diseases and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA. <sup>5</sup>Department of Chemistry, University of Illinois, Urbana, IL 61801, USA. <sup>6</sup>Center for Biophysics and Computational Biology, University of Illinois, Urbana, IL 61801, USA. <sup>7</sup>Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093, USA. <sup>8</sup>Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093, USA.

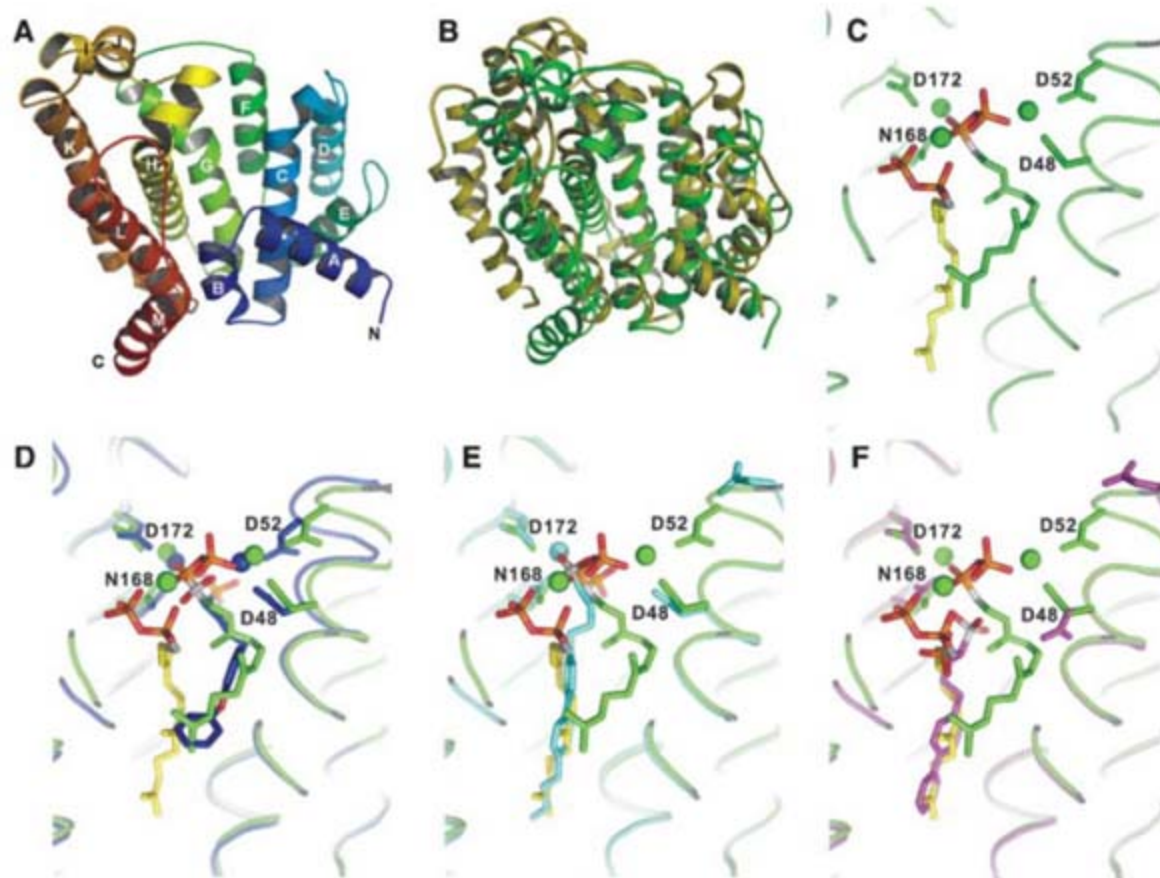
\*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: [vnizet@ucsd.edu](mailto:vnizet@ucsd.edu) (V.N.); [ahjwang@gate.sinica.edu.tw](mailto:ahjwang@gate.sinica.edu.tw) (A.H.-J.W.); [eo@chad.scs.uiuc.edu](mailto:eo@chad.scs.uiuc.edu) (E.O.)



**Fig. 1.** Biosynthetic pathways. **(A)** Staphyloxanthin biosynthesis (in *S. aureus*). **(B)** Cholesterol (in humans) and ergosterol (in, e.g., yeasts and some parasitic protozoa) biosynthesis. Each biosynthetic pathway involves initial formation of presqualene diphosphate, catalyzed by CrtM (*S. aureus*) or by squalene synthase (SQS). In *S. aureus*, the NADPH reduction step is absent, resulting in production of dehydrosqualene, not squalene.

**Fig. 2.** X-ray crystallographic structures. **(A)** X-ray structure of *S. aureus* CrtM. **(B)** Superposition of CrtM and human squalene synthase structures. There is a 5.5 Å  $C_{\alpha}$  RMS deviation between the two structures. **(C)** Close-up view of FsPP bound to CrtM. **(D)** Close-up view of *S. aureus* CrtM with bound BPH-652. **(E)** *S. aureus* CrtM with bound BPH-698. **(F)** *S. aureus* CrtM with bound BPH-700. In **(C)** to **(F)**, the FsPP ligands are in green or yellow; BPH-652, BPH-698, and BPH-700 (and associated  $Mg^{2+}$ ) are in blue, cyan, and magenta, respectively. Key contacts with Asp (D) and Asn (N) residues are indicated.



conventional screening campaigns are not well suited for selecting such inhibitors, because virulence factors typically do not affect bacterial cell growth but rather exert their activity *in vivo*. In *S. aureus*, an important virulence factor is the carotenoid pigment staphyloxanthin. This pigment acts as an antioxidant, with its numerous conjugated double bonds enabling the detoxification of host immune system-generated reactive oxygen species (ROS) such as  $O_2^-$ ,  $H_2O_2$ , and  $HOCl$  (4, 5). Bacteria that lack the carotenoid pigment grow normally, but they are rapidly killed by ROS from host neutrophils and are deficient in skin abscess formation (4). Blocking staphyloxanthin biosynthesis is therefore a potentially attractive therapeutic target (3, 4), and the bright golden coloration of the virulence factor facilitates inhibitor screening.

The first committed step in staphyloxanthin biosynthesis, catalyzed by the *S. aureus* dehydrosqualene synthase (CrtM) enzyme, involves the 1,1' or head-to-head condensation of two molecules of farnesyl diphosphate to produce the  $C_{30}$  species, presqualene diphosphate (5), which then undergoes skeletal rearrangement and further loss of diphosphate to produce dehydrosqualene (Fig. 1A). Successive dehydrogenations yield 4,4'-diaponeurosporene, which is then further oxidized, glycosylated, and esterified to give the carotenoid, staphyloxanthin (Fig. 1A). The structure of dehydrosqualene is very similar to that of the squalene used in cholesterol biosynthesis in humans, and both

dehydrosqualene and squalene biosyntheses proceed through presqualene diphosphate. The pathways diverge at this intermediate, with a reduced nicotinamide adenine dinucleotide phosphate (NADPH)-catalyzed reductive step present in squalene synthesis but not in dehydrosqualene synthesis (Fig. 1B). Thus, *S. aureus* CrtM and human squalene synthase (SQS) might possess structural similarities, although only modest sequence homology (30% identity, 36% similarity) is found by Clustal (6) amino acid alignment (fig. S1). However, human SQS is known to have considerable structural similarity to other prenyl synthase enzymes, including farnesyl diphosphate synthase, pentalenene synthase, and 5-*epi*-aristolochene synthase, despite the lack of sequence homology (7).

To probe this structural question, we cloned, expressed, purified, and crystallized the CrtM protein from *S. aureus* and solved its structure by x-ray crystallography to 1.58 Å. Data collection and refinement statistics for the selenomethionine-

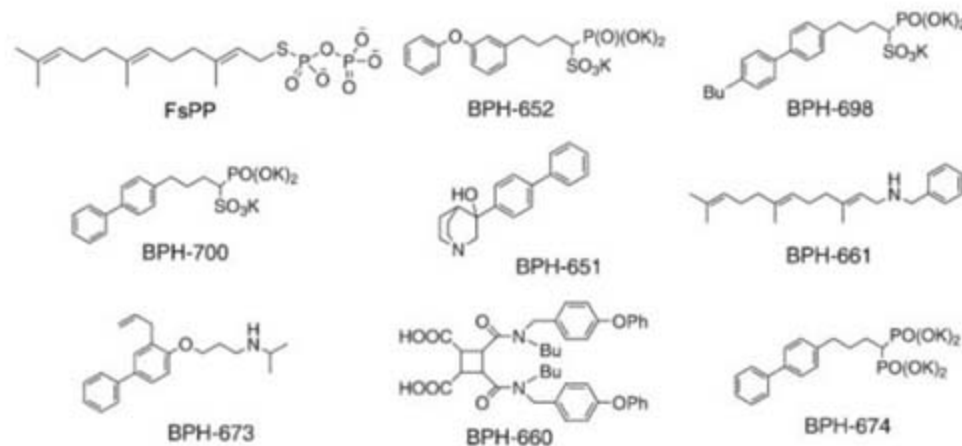
substituted and wild-type proteins are presented in tables S1 and S2. CrtM crystallizes in the  $P3_221$  space group and there is one molecule per asymmetric unit. The overall fold (Fig. 2A) shows clear similarity to that seen in human SQS (PDB accession number 1EZF), as can be seen in the superposition (Fig. 2B), with a 5.5 Å root-mean-square (RMS) deviation between the  $C_\alpha$  atoms in the two structures. CrtM is all helical and has a large central cavity (also seen in other isoprenoid synthases) capable of accommodating the  $C_{30}$  product, dehydrosqualene.

To see how inhibitors might bind to CrtM, we first crystallized the protein with farnesyl thiodiphosphate (FsPP, Fig. 3), a nonreactive analog of farnesyl diphosphate, the substrate for CrtM. FsPP has previously been found to be a substrate-analog inhibitor of other prenyl transferases, including geranylgeranyl diphosphate synthase and undecaprenyl diphosphate synthase (8). However, in each of those struc-

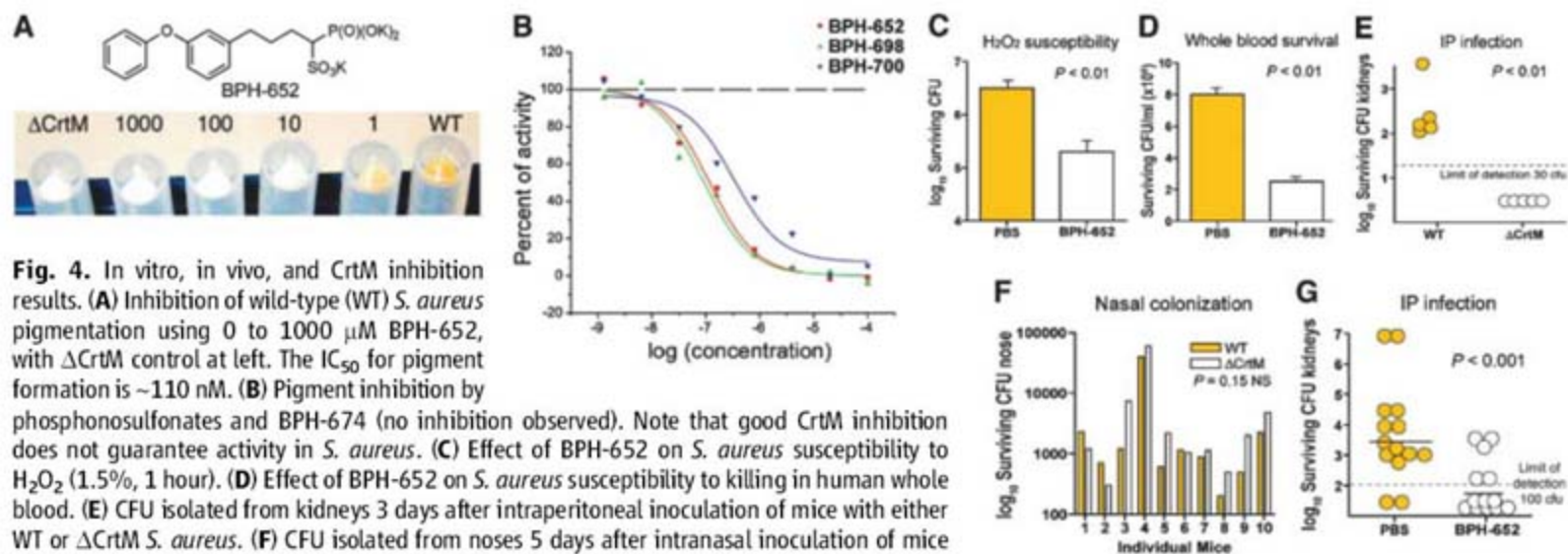
tures, only one FsPP binds per protein. In CrtM, however, two FPP molecules must come together to form dehydrosqualene, and indeed, two FsPP molecules were found in the large central cavity (Fig. 2C). Their diphosphate head groups interact with three  $Mg^{2+}$  ions, which in turn interact with Asp residues in the two conserved Asp-X-X-X-Asp repeats (fig. S1) seen in many prenyl synthases. The space group of FsPP·CrtM is  $P3_121$ , and there are two molecules per asymmetric unit. Electron densities for the two ligands and key protein-ligand interactions are shown in figs. S2 and S3. This ligand binding pattern is likely to define the location of the presqualene diphosphate intermediate binding site.

Structural similarity raised the possibility that human SQS inhibitors developed as potential cholesterol-lowering drugs (9–12) might also be active against CrtM. We synthesized eight such compounds (Fig. 3) and tested them for activity in CrtM inhibition (13). There is considerable structural diversity among these compounds: Some mimic the FPP substrate (e.g., the phosphonosulfonates BPH-652, BPH-698, and BPH-700); others presumably act as analogs of carbocations or transition states (e.g., the amines, BPH-651, BPH-661, and BPH-673); and others, such as the dicarboxylic acid BPH-660, have a less certain mechanism of action.

Of the eight compounds tested, only the three phosphonosulfonates were found to be CrtM inhibitors (BPH-652, inhibitor constant  $K_i = 1.5$  nM; BPH-698,  $K_i = 135$  nM; BPH-700,  $K_i = 6$  nM; fig. S4). Each phosphonosulfonate also had potent activity against *S. aureus* pigment formation in vitro, as shown for BPH-652 in Fig. 4A, with median inhibitory concentration ( $IC_{50}$ ) values in the range 100 to 300 nM (Fig. 4B). The bisphosphonate BPH-674 (Fig. 3) was a more potent CrtM inhibitor ( $K_i = 0.2$  nM; fig. S4) than any of the phosphonosulfonates



**Fig. 3.** Chemical structures of farnesyl thiodiphosphate (FsPP) and eight squalene synthase inhibitors. There are many different types of SQS inhibitor, but only the phosphonosulfonates (and related bisphosphonates) inhibit CrtM. To date, only the phosphonosulfonates have been found to have activity in *S. aureus*.



**Fig. 4.** In vitro, in vivo, and CrtM inhibition results. (A) Inhibition of wild-type (WT) *S. aureus* pigmentation using 0 to 1000  $\mu$ M BPH-652, with  $\Delta$ CrtM control at left. The  $IC_{50}$  for pigment formation is  $\sim$ 110 nM. (B) Pigment inhibition by phosphonosulfonates and BPH-674 (no inhibition observed). Note that good CrtM inhibition does not guarantee activity in *S. aureus*. (C) Effect of BPH-652 on *S. aureus* susceptibility to  $H_2O_2$  (1.5%, 1 hour). (D) Effect of BPH-652 on *S. aureus* susceptibility to killing in human whole blood. (E) CFU isolated from kidneys 3 days after intraperitoneal inoculation of mice with either WT or  $\Delta$ CrtM *S. aureus*. (F) CFU isolated from noses 5 days after intranasal inoculation of mice with 1:1 mixture of WT and  $\Delta$ CrtM *S. aureus*. (G) Effect of BPH-652 on *S. aureus* survival in the kidney after intraperitoneal infection. In vitro data shown are representative of at least three identical experiments. In vivo data in (G) are compiled from two sets of experiments, using different numbers of mice, performed under the same conditions (13).

but had no activity at all in pigment inhibition *in vitro*, due perhaps to poor bacterial cell uptake or increased efflux. None of the four amine/carboxylate human SQS inhibitors had any activity in inhibiting CrtM or pigment formation, although all are known to be potent human SQS inhibitors. Taken together, these results pointed to phosphonosulfonates as being the most promising class of CrtM inhibitors for further investigation, beginning with structure determination.

All three phosphonosulfonates, BPH-652, BPH-698, and BPH-700, yielded well-resolved  $2F_{\text{obs}} - F_{\text{calc}}$  densities (fig. S5); the refined structures obtained are shown in Fig. 2, D to F, superimposed on the FsPP structure. Full crystallographic details are given in table S2, and ligand interaction results are shown in figs. S6 to S8. Interestingly, in all three structures, we found evidence for only one phosphonosulfonate bound per CrtM, not the two ligands that might have been anticipated from the FsPP structure. Moreover, all three inhibitors have different binding modes. BPH-652 (Fig. 2D) binds into the FsPP-1 site with two  $\text{Mg}^{2+}$ , BPH-698 binds into the FsPP-2 site (Fig. 2E) with only one  $\text{Mg}^{2+}$ , and BPH-700 binds into the FsPP-2 site with no  $\text{Mg}^{2+}$  (Fig. 2F). The phosphonosulfonate side chains do, however, closely track the locations of the two FsPP inhibitor side chains. FsPP-1 is highly bent (green structure in Fig. 2C), and BPH-652 is able to track this bent geometry (Fig. 2D), due at least in part to the presence of its central ether O atom. On the other hand, the (linear) biphenyl side chains in the BPH-698 and BPH-700 structures can be closely superimposed on the fully extended famesyl side chain in the FsPP-2 site (yellow in Fig. 2, C, E, and F). Additional views of all superimposed ligands are shown in fig. S9.

Although the aryl side chains in the two biphenyl phosphonosulfonates are closely aligned (Fig. 2, E and F), the head groups bind very differently. With BPH-698, the phosphonosulfonate head group is involved in electrostatic interactions with one  $\text{Mg}^{2+}$ , but in BPH-700 (which lacks the 4'-butyl side chain), no  $\text{Mg}^{2+}$  is found in the structure; instead, the phosphonosulfonate is involved in electrostatic or hydrogen bond interactions with His<sup>18</sup> and Tyr<sup>41</sup>. With all three phosphonosulfonate inhibitors, electrostatic interactions between ligand and protein are important, with eight or nine highly hydrophobic amino acid side chains contacting the inhibitor side chains.

To probe the possible use of a phosphonosulfonate CrtM inhibitor as an antimicrobial, we performed a series of experiments to determine whether such a compound was (i) nontoxic to mammalian cells, (ii) able to restore *S. aureus* killing by oxidants and phagocytes, and (iii) able to limit *S. aureus* disease progression *in vivo*. BPH-652 was selected for these experiments because it has a good  $\text{IC}_{50}$  in pigment inhibition (110 nM; Fig. 4B) and because *S*-BPH-652

(and its analogs) have been advanced through preclinical testing in rats and in two human clinical trials (14, 15) as cholesterol-lowering agents.

As expected, BPH-652 had no effect on the growth of three human cell lines (MCF-7, NCI-H460, and SF-268), because only cholesterol biosynthesis is targeted and cholesterol is generally abundant in serum (or diet). This indicates low toxicity, consistent with the results of the clinical trials (on *S*-BPH-652). Also, incubation with up to 2 mM BPH-652 affected neither *S. aureus* growth characteristics nor survival through 48 hours in culture. However, after incubation with 100  $\mu\text{M}$  BPH-652, the resulting nonpigmented (white) *S. aureus* were more susceptible to killing by 1.5% hydrogen peroxide by a factor of ~15, and, relative to normally pigmented *S. aureus* treated with phosphate-buffered saline (PBS) control, were less able to survive in freshly isolated human whole blood by a factor of ~4 (Fig. 4, C and D), as expected because they contained no carotenoid pigment to act as an antioxidant.

Finally, we sought to establish whether inhibition of staphyloxanthin biosynthesis by BPH-652 could represent a purely virulence factor-based therapy against *S. aureus* infection. We first extended our previous observations on the contribution of staphyloxanthin to abscess formation (4) to a systemic *S. aureus* infection model. Staphyloxanthin was found to promote invasive disease potential, because mice inoculated with  $10^8$  colony-forming units (CFU) of wild-type *S. aureus* (by intraperitoneal injection) developed a sustained infection, with bacteria recovered from the kidneys 72 hours later; an isogenic *S. aureus* mutant lacking the CrtM enzyme was cleared by the host after similar challenge (Fig. 4E). Conversely, the contribution of *S. aureus* pigmentation to mucosal colonization after intranasal inoculation was negligible (Fig. 4F). These data are consistent with a primary function of staphyloxanthin in resisting the host oxidant-based phagocytic defenses that are present in disease contexts but are absent on mucosal surfaces without inflammation.

Under the same intraperitoneal challenge used in Fig. 4E, we then treated one group of mice ( $n = 14$ ) with 0.5 mg of BPH-652 twice per day (days -1, 0, 1, and 2), and a second group ( $n = 13$ ) with equivalent volume injections of PBS control. Upon killing at 72 hours, *S. aureus* bacterial counts in the kidneys of the mice treated with BPH-652 were significantly lower than those of the control group ( $P < 0.001$ ), with 8 of 13 below the detection threshold, versus only 2 of 14 in the control group (Fig. 4G); on average, this result corresponds to a 98% decrease in surviving bacteria in the treatment group.

Thus, the CrtM (dehydroisqualene synthase) enzyme from *S. aureus* is a target for anti-

infective therapy based on virulence factor neutralization, and a drug candidate already tested in humans in the context of cholesterol-lowering therapy provides a good lead. These results therefore provide a basis for rational drug design against this major human pathogen, and provide proof of principle of the utility of an anti-infective drug without direct bactericidal properties that instead renders the pathogen susceptible to normal host innate immune clearance. Our approach, and other virulence factor-based concepts (16, 17) for highly specific antistaphylococcal therapy, also offer theoretical advantages for reducing selection pressure toward the emergence of resistance, both in the pathogen and in our normal commensal microflora (3, 18).

## References and Notes

1. E. A. Bancroft, *JAMA* **298**, 1803 (2007).
2. R. M. Kleven et al., *JAMA* **298**, 1763 (2007).
3. National Research Council, in *Treating Infectious Diseases in a Microbial World: Report of Two Workshops on Novel Antimicrobial Therapeutics* (National Academies Press, Washington, DC, 2006), pp. 21–22.
4. G. Y. Liu et al., *J. Exp. Med.* **202**, 209 (2005).
5. A. Clauditz, A. Resch, K. P. Wieland, A. Peschel, F. Gotz, *Infect. Immun.* **74**, 4950 (2006).
6. J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* **22**, 4673 (1994).
7. J. Pandit et al., *J. Biol. Chem.* **275**, 30610 (2000).
8. R. T. Guo et al., *Proc. Natl. Acad. Sci. U.S.A.* **104**, 10022 (2007).
9. H. Hiyoshi et al., *J. Lipid Res.* **41**, 1136 (2000).
10. T. Nishimoto et al., *Br. J. Pharmacol.* **139**, 911 (2003).
11. T. Ishihara et al., *Bioorg. Med. Chem.* **11**, 2403 (2003).
12. D. R. Magnin et al., *J. Med. Chem.* **39**, 657 (1996).
13. See supporting material on Science Online.
14. A. Sharma, P. H. Slugg, J. L. Hammett, W. J. Jusko, *Pharm. Res.* **15**, 1782 (1998).
15. A. Sharma, P. H. Slugg, J. L. Hammett, W. J. Jusko, *J. Clin. Pharmacol.* **38**, 1116 (1998).
16. H. Ton-That, O. Schneewind, *J. Biol. Chem.* **274**, 24316 (1999).
17. G. J. Lyon, P. Mayville, T. W. Muir, R. P. Novick, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13330 (2000).
18. L. E. Alksne, S. J. Projan, *Curr. Opin. Biotechnol.* **11**, 625 (2000).
19. Supported by U.S. Public Health Service grants AI07482 (G.Y.L.), HD051796 (V.N.), and GM073216 and GM65307 (E.O.); Academia Sinica and National Core Facility of High-Throughput Protein Crystallography grant NSC95-3112-B-001-015-Y (A.H.-J.W.); a Burroughs-Wellcome Career Award (G.Y.L.); and a Leukemia and Lymphoma Society Special Fellowship (Y.S.). We thank K. Krysiak for carrying out the mammalian cell growth inhibition assays. Diffraction data were obtained at the National Synchrotron Radiation Research Center of Taiwan, SPring-8 and the Photon Factory in Japan, and at the Advanced Light Source, Berkeley, CA. The coordinates for the CrtM, FsPP, BPH-652, BPH-698, and BPH-700 structures have been deposited in the Protein Data Bank under accession numbers 2ZCO, 2ZCP, 2ZCQ, 2ZCR, and 2ZCS, respectively.

## Supporting Online Material

[www.sciencemag.org/cgi/content/full/1153018/DC1](http://www.sciencemag.org/cgi/content/full/1153018/DC1)

Materials and Methods

Figs. S1 to S9

Tables S1 and S2

References

15 November 2007; accepted 28 January 2008

Published online 14 February 2008;

10.1126/science.1153018

Include this information when citing this paper.

# High-Resolution Mapping of Crossovers Reveals Extensive Variation in Fine-Scale Recombination Patterns Among Humans

Graham Coop,<sup>1\*</sup> Xiaoquan Wen,<sup>1</sup> Carole Ober,<sup>1,2</sup> Jonathan K. Pritchard,<sup>1\*</sup> Molly Przeworski<sup>1\*</sup>

Recombination plays a crucial role in meiosis, ensuring the proper segregation of chromosomes. Recent linkage disequilibrium (LD) and sperm-typing studies suggest that recombination rates vary tremendously across the human genome, with most events occurring in narrow “hotspots.” To examine variation in fine-scale recombination patterns among individuals, we used dense, genome-wide single-nucleotide polymorphism data collected in nuclear families to localize crossovers with high spatial resolution. This analysis revealed that overall recombination hotspot usage is similar in males and females, with individual hotspots often active in both sexes. Across the genome, roughly 60% of crossovers occurred in hotspots inferred from LD studies. Notably, however, we found extensive and heritable variation among both males and females in the proportion of crossovers occurring in these hotspots.

Errors in the recombination process during meiosis underlie a variety of chromosomal abnormalities and greatly increase the risk of nondisjunction (1–3). Nonetheless, the total number of recombination events varies significantly among individuals (4), and rates of genetic exchange over fine scales are known to differ in males (5–8). These observations hint at extensive variation in many aspects of the recombination process (9), the nature and extent of which have yet to be systematically characterized. In particular, because observations of

recombination between closely linked markers come from sperm typing, we still know little about fine-scale patterns of recombination in females.

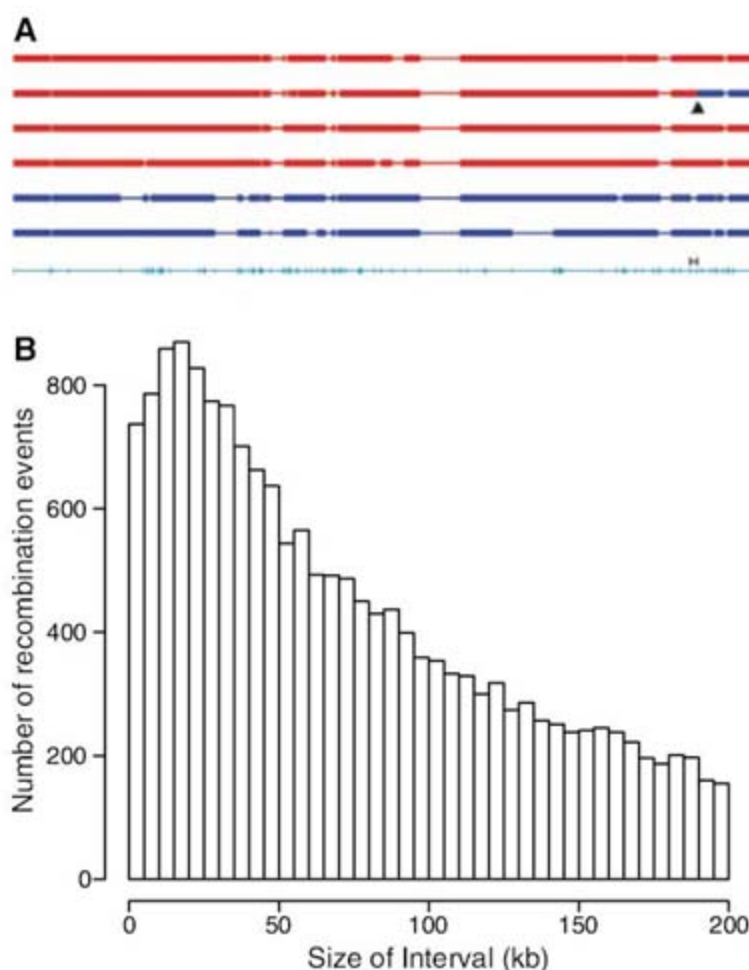
With the recent advent of high-density genotyping platforms, it is now feasible to study fine-scale patterns of recombination with pedigree data. To test this approach, we analyzed genome-wide single-nucleotide polymorphism data from the Affymetrix GeneChip Mapping 500K Array Set (Affymetrix, Santa Clara, CA) in 725 related Hutterites, a population of European descent

(10). The 725 individuals form part of a known, 1650-person, 13-generation pedigree, which we broke down into a set of 82 overlapping nuclear families for purposes of analysis. Of the 82 families, 50 have between four and ten genotyped children, 18 have three, and 14 have two, allowing us to infer recombination events in a total of 364 male and 364 female gametes. Although the number of meioses is smaller than that of Kong *et al.* (11) (728 versus 1257 meioses), our marker density is nearly 100-fold higher, allowing us to define crossover locations with high spatial resolution.

To infer recombination events from genotype data in nuclear families with two or more children, we devised an algorithm that effectively phases the parental chromosomes and identifies positions where a child’s chromosome switches from copying one parental haplotype to the other [Fig. 1A and supporting online material (SOM) text]. This approach identified 24,095 autosomal crossovers in 728 meioses, of which 12,278 were localized to an interval of less than 100 kb and 4,854 to within 30 kb (Fig. 1B). We inferred a mean of 39.6 recombination events per gamete [95% confidence interval (CI):

<sup>1</sup>Department of Human Genetics, University of Chicago, 920 East 58th Street, Cummings Life Science Center, Chicago, IL 60637, USA. <sup>2</sup>Department of Obstetrics and Gynecology, University of Chicago, Chicago, IL 60637, USA.

\*To whom correspondence should be addressed. E-mail: gcoop@uchicago.edu (G.C.); pritch@uchicago.edu (J.K.P.); mfp@uchicago.edu (M.P.)



**Fig. 1.** (A) Transmissions of a 4-megabase (Mb) region on chromosome 5 from a mother to her six children. The blue hashes on the lowest line indicate the location of informative markers in the mother, whereas the six blue and red lines above label the two estimated maternal haplotypes, with the thinner sections indicating the missing data. The triangle points to the inferred recombination event. (B) A histogram of interval sizes of recombination events resolved to within 200 kb (representing ~70% of the total), within which we inferred crossovers to have occurred.



38.5 to 40.6] in females and a mean of 26.2 recombination events per gamete (95% CI: 25.6 to 26.7) in males. Both these estimates, as well as our recombination rate estimates at the megabase scale, agree closely with those of previous studies (11) (SOM text). This agreement implies that our algorithm calls recombination events reliably and that, at least at this scale of comparison, the Hutterite and Icelandic populations have similar overall recombination rates.

We then examined recombination among individuals, confirming the existence of significant variation ( $SD = 4.71$ ,  $P = 0.0007$ ) in the mean number of recombination events among females (11, 12). A previous report (13) had suggested that mothers with higher recombination rates have slightly more offspring and that viable offspring of older mothers tend to have higher recombination rates (13). We saw a similar effect in the Hutterites: We estimated that children born to mothers aged 35 years or older have, on average, an extra 3.1 maternal recombination events as compared with those born to mothers below the age of 35 (one-sided  $P = 0.028$ , from a stringent within-family permutation test). This maternal age effect may reflect selection against oocytes that have insufficient numbers of recombination events to overcome insults to the meiotic system accumulated over time (13, 14).

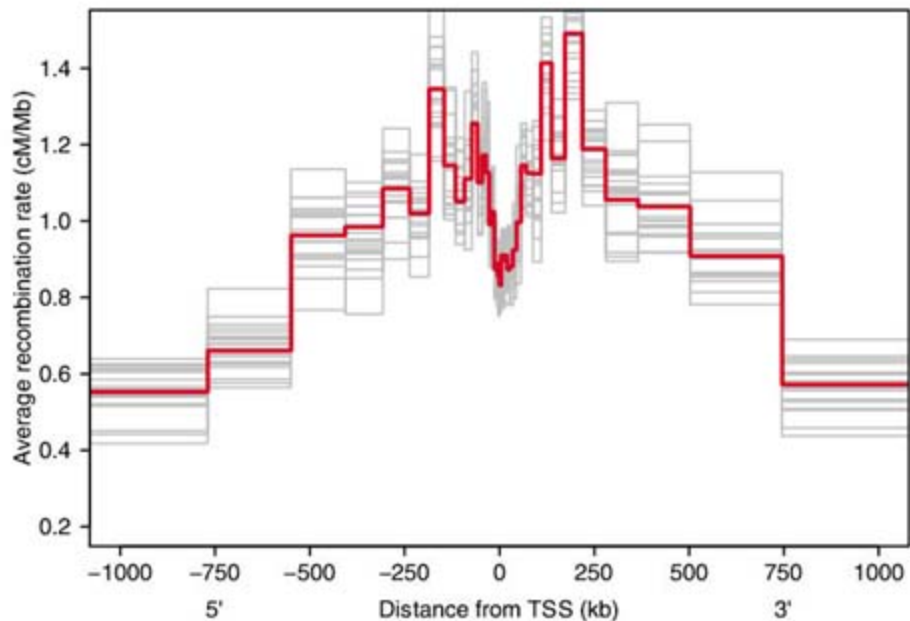
Among males, we also found significant variation in the total number of recombination events ( $SD = 2.59$ ,  $P = 0.0001$ ), as reported for cytogenetic studies (15) but, until recently (16), not seen in pedigree studies. Moreover, in males, we detected significant variation in the number of crossovers on individual chromosomes (especially chromosome 19), even after correcting for genome-wide recombination (SOM text). Notably, the chromosomes with significant variation in males show no such evidence in females, suggesting that there may be sex-specific modifiers of recombination rate at this scale (9, 17). Unlike in females, we did not detect an effect of paternal age on recombination.

On a broad scale, recombination rates are known to increase with gene density (11), which is consistent with a link between transcription and recombination, as found in yeast (18). At a finer scale, however, patterns of LD suggest that recombination rates are actually reduced near genes and highest at a short distance from the start positions of genes (19). Because LD patterns are shaped not only by recombination but also by natural selection, the interpretation of this finding is not clear-cut; indeed, the observation of increased LD within genes has been interpreted both as a signal of natural selection in genes (20) and as evidence for reduced recombination (19).

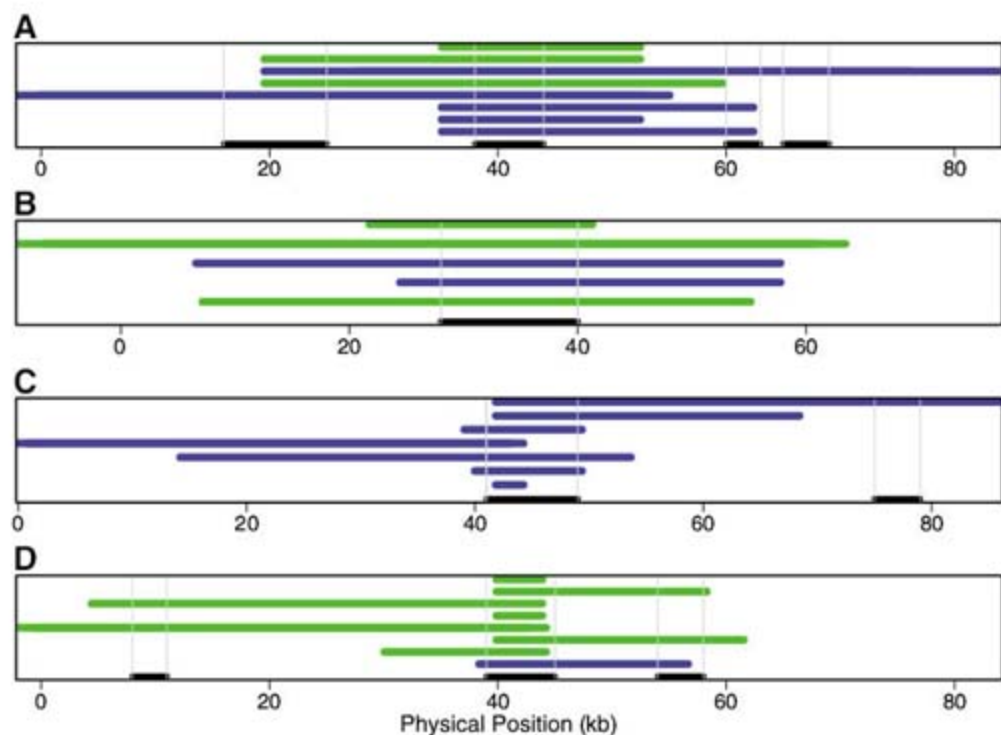
To resolve this issue using directly observed recombination events, we estimated the average recombination rates as a function of distance from the nearest transcription start site (TSS) (Fig. 2 and SOM text). Of the 4854 recombination events that were refined to within 30 kb, we

found that recombination rates are typically low near the TSS (both upstream and downstream) and are highest in regions tens or hundreds of kilobases from the nearest TSS. These results indicate that recombination tends to occur in more-

distant intergenic sites that may be less likely to be associated with promoter function, implying that the primary cause of increased LD near the TSS is decreased recombination rather than selection.



**Fig. 2.** Distribution of recombination relative to genes. The red line plots the estimated, average recombination rate as a function of distance from the nearest TSS, calculated with recombination events refined to within 30 kb. The physical length of each bin is indicated by the length of the horizontal line. The 20 gray lines show averages calculated from bootstrap resampling of recombination events, as a measure of the uncertainty in our estimates. cM, centimorgan.



**Fig. 3.** Overlap of recombination events with four specific hotspots inferred from LD analyses. These regions were chosen because they contain some of the most active hotspots seen over all chromosomes (SOM text). Each panel displays results for one region, with the physical position (in kilobases) denoted on the x axis. Only recombination events localized to within 100 kb are shown, with intervals containing male and female crossovers indicated in green and blue, respectively. The locations of the hotspots estimated from LD data are shown along the bottom of each panel as black lines, and vertical light gray lines indicate their boundaries. (A) Region located at 69.5 kb on chromosome 17. (B) Region located at 58.7 kb on chromosome 19. (C) Region located at 119.5 kb on chromosome 10. (D) Region located at 132.5 kb on chromosome 11.

Sperm-typing and LD analyses suggest that most (60 to 70%) crossover events occur in about 10% of the genome (19), a heterogeneity largely due to 1- to 2-kb hotspot regions that experience sharply elevated recombination relative to that of the background (8, 9, 19, 21–23). Although such studies have vastly improved our knowledge of fine-scale rates, sperm-typing studies are labor-intensive and only informative about male rates. In turn, LD-based estimates rely on a simple population genetic model and are both sex-averaged (over both male and female ancestors of the sample) and time-averaged (over many ancestral generations); consequently, such estimates cannot be used to learn about variation in rates among individuals. In contrast, our high-resolution pedigree data allow us to directly observe crossover events in transmissions from both males and females and to examine inter-individual variation.

To learn more about the nature of hotspots, we considered all recombination events in our data whose location could be inferred to within 30 kb (2910 female and 1944 male events; see SOM for results with other cutoffs). To assess the congruence between LD- and pedigree-based estimates of recombination, we examined how often these well-resolved recombination events overlapped with 32,996 putative hotspots estimated from LD patterns in the Phase II HapMap data (24) (SOM text). We found that 72% of crossovers overlap a hotspot, when just 32% would be expected to do so by chance. We then used a likelihood method to estimate the true proportion of recombination events that takes place in hotspots, accounting for the possibility that an event overlaps simply by chance (SOM text). We found that 60% (95% CI: 58 to 61%) of recombination events occurred in hotspots, which

closely agrees with analyses of LD data (19). A number of the LD hotspots that were overlapped by our inferred recombination events appear to be extremely active: For example, three of the hotspots shown in Fig. 3 are potentially active in as many as 1% of meioses (see SOM text).

Overall, our results support the picture of recombination rate heterogeneity as suggested by LD analyses, notably in terms of the fraction of crossovers occurring in hotspots. This concordance implies that hotspots detected in extant populations have persisted for at least the time scale detectable in LD (i.e., thousands of generations). Our findings do not mean that every inference of a hotspot from LD data is true; our well-resolved recombination events only overlap a total of 3200 hotspots, leaving many predicted hotspots to be confirmed.

At broad scales, females and males are known to differ dramatically in their recombination rates (11, 12), whereas, at finer scales, very little is known about differences between sexes. One hint that recombination rate heterogeneity may be similar between the two sexes is that LD data from the X chromosome—which (outside of the pseudoautosomal regions) recombines only in females—show patterns of hotspots that are roughly similar to those on autosomes (19). Our data show that indeed overall hotspot use is quite similar in the two sexes. Across the genome, the fraction of crossovers that occur in recombination hotspots inferred from LD differs only slightly between males (62%; 95% CI: 59 to 64%) and females (57%; 95% CI: 55 to 59%). Moreover, inspection of specific hotspots revealed that they are often active in both males and females, as they coincide with well-resolved recombination events in both sexes (Fig. 3, A and B). A subset of hotspots, however, seems to be used mainly by

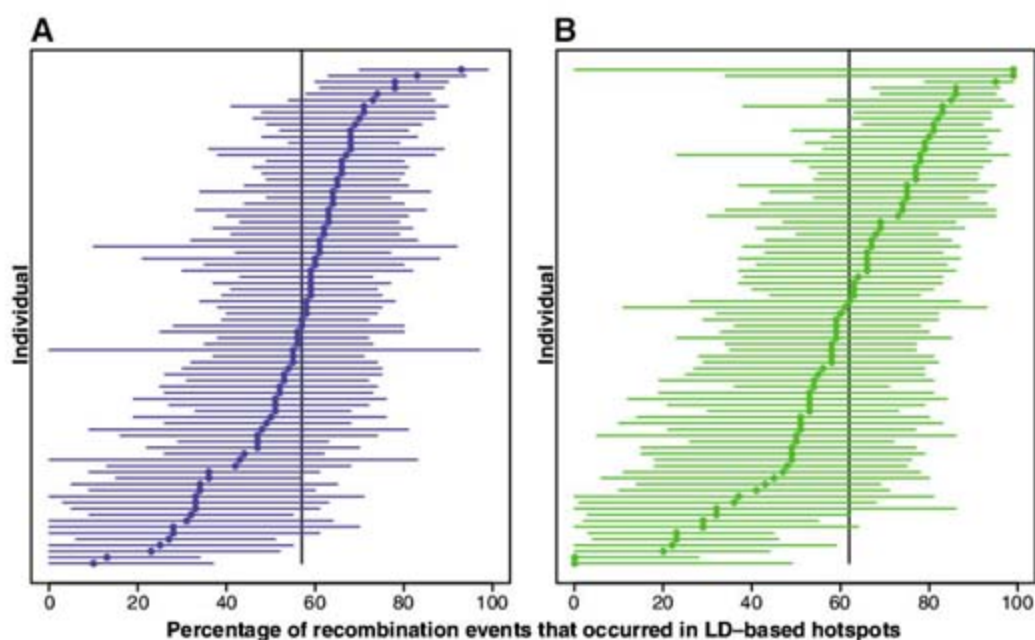
one sex or the other. For example, the hotspot in Fig. 3C is potentially active only in females, whereas the hotspot in Fig. 3D appears to be active mostly in males (SOM text). Our analyses indicated that the sex-specific use of individual hotspots is explained in part by differences in broad (megabase)-scale rates but that there is also considerable variation between sexes below the megabase scale. We also examined the overlap between recombination events within and between sexes, controlling for the broader-scale rate (for details, see the SOM). Together, these analyses suggest that males may use a smaller subset of hotspots than females.

Although we found no marked difference in the average hotspot use between sexes, we noted extensive variation among both males and females in the fraction of crossovers that occur in hotspots. For each parent, averaging across all their offspring, we estimated the genome-wide proportion of events that occur in LD-based hotspots ( $\alpha$ ) (Fig. 4). The variation in  $\alpha$  among individuals is highly significant by a likelihood ratio test ( $P$  value from permutation test:  $P < 0.002$ , for both sexes). Moreover, the narrow-sense heritability of the fraction of crossovers in LD-based hotspots is estimated to be 0.22, which is significantly larger than 0 ( $P = 0.01$ , with a test that accounts for relatedness across the entire Hutterite pedigree) (25). Thus, genome-wide use of LD-based hotspots is significantly variable among individuals (males and females), and this variation is heritable.

One interpretation of this finding is that some individuals use recombination hotspots less frequently than others. However, because hotspots detected in LD data are likely to have been active for thousands of years, it may be that all individuals use hotspots equally, but some tend to use newer or weaker hotspots that are less likely to be found in analyses of LD. Regardless of the interpretation, the finding of heritable variation in LD-based hotspot use points to heritable differences among individuals in some aspect of the recombination machinery.

This result is particularly interesting in light of recent reports suggesting that hotspot locations have evolved rapidly since the split between humans and chimpanzees (26–29), because differences in trans-acting factors in humans and chimpanzees could account for the marked difference in hotspot locations between the two species. Moreover, our finding offers a possible solution to the hotspot paradox (i.e., the existence of hotspots despite biased gene conversion against alleles that promote them) (6, 7, 30). A single change in the recombination machinery could create many new hotspots in the genome, counteracting the removal of individual hotspots from the population by biased gene conversion (9, 31).

These analyses uncovered tremendous variation in recombination rates over all genomic scales considered and, in particular, heritable variation in hotspot use. It should now be possi-



**Fig. 4.** The percentage of crossovers inferred to have occurred in LD-based hotspots in each individual. The maximum likelihood estimate (MLE) for each individual [females in (A) and males in (B)] is shown as a circle, and the 95% CIs are indicated by the length of the horizontal lines. Individuals are ordered by their MLE. The black vertical line in each panel shows the overall MLE.

ble to map the genetic basis for variation in different aspects of the recombination process, with high-density genotyping data. Identifying the loci that contribute to this variation will offer unparalleled insights into the genetic basis of recombination rate variation and the selective forces governing the evolution of recombination rates (9).

#### References and Notes

1. T. Hassold, P. Hunt, *Nat. Rev. Genet.* **2**, 280 (2001).
2. K. Inoue, J. R. Lupski, *Annu. Rev. Genomics Hum. Genet.* **3**, 199 (2002).
3. J. A. Bailey, E. E. Eichler, *Nat. Rev. Genet.* **7**, 552 (2006).
4. A. Lynn, T. Ashley, T. Hassold, *Annu. Rev. Genomics Hum. Genet.* **5**, 317 (2004).
5. J. Yu et al., *Am. J. Hum. Genet.* **59**, 1186 (1996).
6. A. J. Jeffreys, R. Neumann, *Nat. Genet.* **31**, 267 (2002).
7. A. J. Jeffreys, R. Neumann, *Hum. Mol. Genet.* **14**, 2277 (2005).
8. I. Tiemann-Boege, P. Calabrese, D. M. Cochran, R. Sokol, N. Arnheim, *PLoS Genet.* **2**, e70 (2006).
9. G. Coop, M. Przeworski, *Nat. Rev. Genet.* **8**, 23 (2007).
10. C. Ober, M. Abney, M. S. McPeck, *Am. J. Hum. Genet.* **69**, 1068 (2001).
11. A. Kong et al., *Nat. Genet.* **31**, 241 (2002).
12. K. W. Broman, J. C. Murray, V. C. Sheffield, R. L. White, J. L. Weber, *Am. J. Hum. Genet.* **63**, 861 (1998).
13. A. Kong et al., *Nat. Genet.* **36**, 1203 (2004).
14. R. U. Vallente, E. Y. Cheng, T. J. Hassold, *Chromosoma* **115**, 241 (2006).
15. T. Hassold et al., *Cytogenet. Genome Res.* **107**, 249 (2004).
16. V. G. Cheung, J. T. Burdick, D. Hirschmann, M. Morley, *Am. J. Hum. Genet.* **80**, 526 (2007).
17. T. Lenormand, *Genetics* **163**, 811 (2003).
18. T. D. Petes, *Nat. Rev. Genet.* **2**, 360 (2001).
19. S. Myers, L. Bottolo, C. Freeman, G. McVean, P. Donnelly, *Science* **310**, 321 (2005).
20. D. A. Hinds et al., *Science* **307**, 1072 (2005).
21. L. Kauppi, A. J. Jeffreys, S. Keeney, *Nat. Rev. Genet.* **5**, 413 (2004).
22. D. C. Crawford et al., *Nat. Genet.* **36**, 700 (2004).
23. G. A. T. McVean et al., *Science* **304**, 581 (2004).
24. International HapMap Consortium, *Nature* **449**, 851 (2007).
25. M. Abney, M. S. McPeck, C. Ober, *Am. J. Hum. Genet.* **66**, 629 (2000).
26. J. D. Wall, L. A. Frisse, R. R. Hudson, A. Di Rienzo, *Am. J. Hum. Genet.* **73**, 1330 (2003).
27. S. E. Ptak et al., *PLoS Biol.* **2**, e155 (2004).
28. S. E. Ptak et al., *Nat. Genet.* **37**, 429 (2005).
29. W. Winckler et al., *Science* **308**, 107 (2005).
30. A. Boulton, R. S. Myers, R. J. Redfield, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 8058 (1997).
31. G. Coop, S. R. Myers, *PLoS Genet.* **3**, e35 (2007).
32. We thank S. Myers and members of the Pritchard, Przeworski, and Stephens labs for discussion. This work was supported by NIH grants GM83098 to M.P.; HG002772 to J.K.P.; and HD21244, HL56399, and HL56533 to C.O.

#### Supporting Online Material

[www.sciencemag.org/cgi/content/full/1151851/DC1](http://www.sciencemag.org/cgi/content/full/1151851/DC1)

SOM Text

Figs. S1 to S10

Tables S1 to S9

References

17 October 2007; accepted 23 January 2008

Published online 31 January 2008;

10.1126/science.1151851

Include this information when citing this paper.

## Sequence Variants in the *RNF212* Gene Associate with Genome-Wide Recombination Rate

Augustine Kong,\* Gudmar Thorleifsson, Hreinn Stefansson, Gisli Masson, Agnar Helgason, Daniel F. Gudbjartsson, Gudrun M. Jonsdottir, Sigurjon A. Gudjonsson, Sverrir Sverrisson, Theodora Thorlacius, Aslaug Jonasdottir, Gudmundur A. Hardarson, Stefan T. Palsson, Michael L. Frigge, Jeffrey R. Gulcher, Unnur Thorsteinsdottir, Kari Stefansson\*

The genome-wide recombination rate varies between individuals, but the mechanism controlling this variation in humans has remained elusive. A genome-wide search identified sequence variants in the 4p16.3 region correlated with recombination rate in both males and females. These variants are located in the *RNF212* gene, a putative ortholog of the *ZHP-3* gene that is essential for recombinations and chiasma formation in *Caenorhabditis elegans*. It is noteworthy that the haplotype formed by two single-nucleotide polymorphisms (SNPs) associated with the highest recombination rate in males is associated with a low recombination rate in females. Consequently, if the frequency of the haplotype changes, the average recombination rate will increase for one sex and decrease for the other, but the sex-averaged recombination rate of the population can stay relatively constant.

Recombination generates part of the diversity that fuels evolution. In humans, it has been suggested that recombination rate must be highly regulated (1), as too little recombination can lead to inaccurate disjunction and aneuploidy (2, 3), whereas ectopic exchange can lead to chromosomal rearrangements (4). Some regions in the genome, known as hotspots, have much higher recombination rate per physical distance unit than the genome as a whole. By using high-density single-nucleotide polymorphism (SNP) data, from which historical recom-

binational events can be inferred, and sperm data, substantial advances have been made in the understanding of local recombination rate (5–11). Furthermore, male and female recombination patterns are different in both genome-wide and regional recombination rates (12, 13). It is also firmly established that genome-wide recombination rate varies substantially among women (12, 13), and there have been hints that this also is true in men (14–16).

Previously, we genotyped a large number of families with a genome-wide microsatellite set of ~1000 markers. This work allowed us to estimate the recombination rate for thousands of men and women and demonstrated that maternal recombination rate increases with the age of the mother

and that there is a positive correlation between the number of children and the recombination rate of a woman (17). A common inversion on chromosome 17q21.31 was also identified that associates with recombination rate and fertility of women (18). Here, we performed a genome-wide scan for variants associated with recombination rate by genotyping with the Illumina Hap300 chip 1887 males and 1702 females with recombination rate estimates [see (19) for a description of study groups]. After quality filtering, 309,241 SNPs were tested for association with recombination frequencies. Male and female recombination rates were studied separately with weighted regression where the weight of a person was proportional to the number of children used to estimate recombination rate. We fitted an additive model with the estimated recombination rate regressed on the number of an allele (0, 1, or 2) a person carried. The results were then adjusted for relatedness between individuals and potential population stratification with the method of genomic control (20). Specifically, standard errors of the effect estimates resulting from the regressions were multiplied by a factor of 1.041 for males and 1.067 for females corresponding to dividing the chi-square test statistics by an adjustment factor of  $1.084 = 1.041^2$  and  $1.138 = 1.067^2$  [see (19) for quality control and statistical analysis].

For the recombination rate of males, three SNPs achieved genome-wide significance ( $P < 1.6 \times 10^{-7}$ , fig. S1). They were rs3796619 ( $P = 1.1 \times 10^{-14}$ ), rs1670533 ( $P = 1.8 \times 10^{-11}$ ), and rs2045065 ( $P = 1.6 \times 10^{-11}$ ), which were all located within a small region in strong linkage disequilibrium (LD) on chromosome 4p16.3 (Fig. 1). The same three SNPs were also associated with the female recombination rate (Table 1). The last two SNPs achieved genome-wide significance; no

deCODE Genetics Inc, 101 Reykjavik, Iceland.

\*To whom correspondence should be addressed. E-mail: [kong@decode.is](mailto:kong@decode.is) (A.K.); [kstefans@decode.is](mailto:kstefans@decode.is) (K.S.)

other SNP did. However, when we compared males with females, the three SNPs each showed opposite effects. For example, allele T of rs3796619 was associated with low recombination rate in males, but high recombination rate in females. These three SNPs were genotyped in a replication sample of 1248 males and 1663 females for whom recombination rate estimates were also available. Overall, 3135 males and 3365 females were genotyped for the three SNPs. A total of 4388 nuclear families and 19,578 individuals genotyped with genome-wide microsatellite

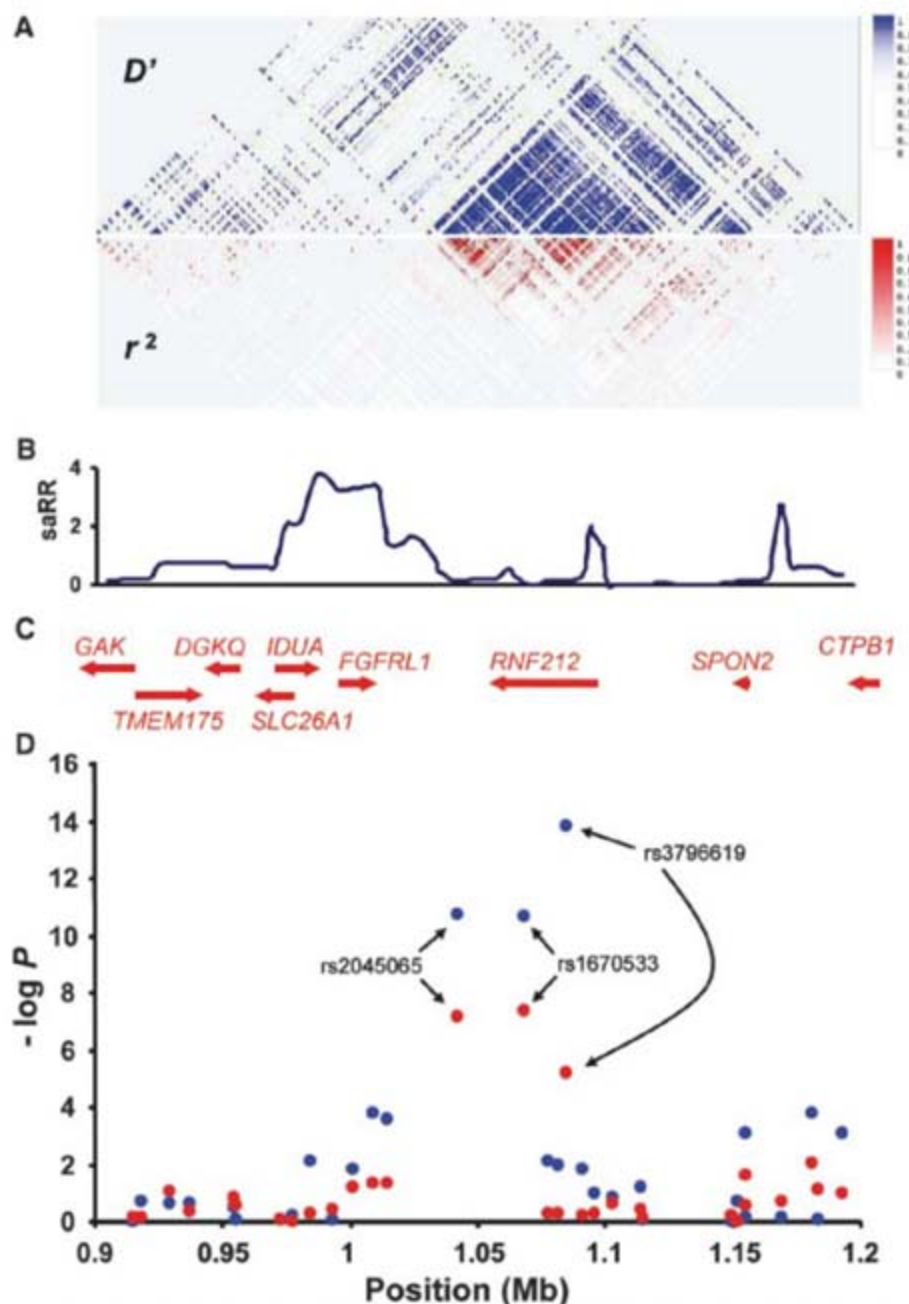
markers (19) contributed to the final recombination rate estimates confirming the original associations. With the samples combined, rs3796619 showed the strongest association with male recombination rate ( $P = 3.2 \times 10^{-24}$ ), and each copy of allele T (compared with C) was estimated to decrease recombination rate by 70.7 centimorgans (cM) (Table 1). For female recombination rate, the SNP rs1670533 showed the strongest association ( $P = 1.9 \times 10^{-12}$ ), and relative to the TT homozygote, each copy of allele C was estimated to increase recombination

rate by 88.2 cM. The third SNP, rs2045065, was highly correlated with rs1670533 ( $r^2 = 0.99$ ), and their effects could not be distinguished from each other. Henceforth, for simplicity, we focused on the joint effects of rs3796619 and rs1670533.

Although rs3796619 and rs1670533 are not surrogates of each other, nonetheless, they are in strong LD ( $D' = 1$ ,  $r^2 = 0.60$ ), and only three of the four possible haplotypes were observed in our sample. Pairwise comparisons of the effects of these three haplotypes on recombination rate were made; an additive model was assumed (Fig. 2, A and B). For males, haplotype [C,T] was associated with significantly higher recombination rate than both [T,T] and [T,C] ( $P = 3.2 \times 10^{-7}$  and  $7.0 \times 10^{-23}$ , respectively). Haplotype [T,T] was associated with a slightly higher recombination rate than haplotype [T,C], but the difference did not reach significance ( $P = 0.061$ ). Hence, most, if not all, of the difference in male recombination rate between the three haplotypes could be explained by SNP rs3796619. For females, haplotype [T,C] was associated with significantly higher recombination rate than both haplotype [T,T] ( $P = 6.6 \times 10^{-7}$ ) and haplotype [C,T] ( $P = 5.4 \times 10^{-11}$ ), although there was not a significant difference between the latter two haplotypes ( $P = 0.47$ ). Hence, the SNP rs1670533 alone may account for all the differences in female recombination rate between these haplotypes.

For the association of rs3796619 with male recombination rate and the association of rs1670533 with female recombination rate, testing the additive model against the full model and treating the genotype as a categorical variable showed that, in both cases, these results were not significantly different from that expected under the additive model ( $P = 0.77$  and  $0.68$ , respectively). It is noteworthy that haplotype [C,T] was associated with high male recombination rate and low female recombination and that the exact opposite was true for haplotype [T,C]. To ensure that these observations were not an artifact of estimating recombination rates through family linkage data, however unlikely, we analyzed the data of 2152 couples, a subset of the 3135 males and 3365 females studied. When the genotypes of the spouses were included in the regression, they were not statistically significant and did not affect the association between the genotypes of the individuals and their estimated recombination rates.

Association results for individual chromosomes, with the estimated effect presented as a percentage of the average recombination rate of a chromosome, are displayed in Fig. 3 (see fig. S2 for effects in centimorgans). For males, allele T of rs3796619 was estimated to have a negative effect on recombination rate for 21 of the 22 autosomes, and this effect was significant ( $P < 0.05$ ) for 13 chromosomes. Although this supports the hypothesis that the effect of the variant is indeed genome-wide, a test of heterogeneity indicated that the percentage change is not the same for all chromosomes ( $P = 0.001$ ). For females, allele C of rs1670533 was estimated to have a positive

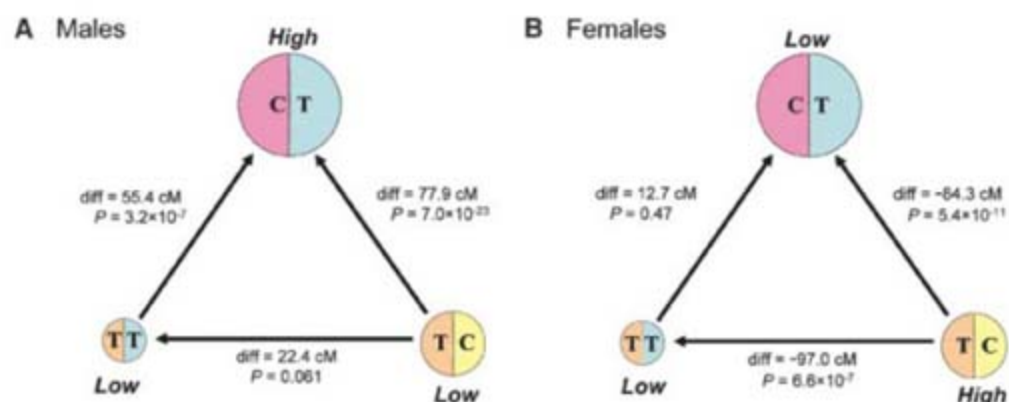


**Fig. 1.** (A) The pairwise correlation structure in a 300-kb interval (0.9 to 1.2 Mb, National Center for Biotechnology Information Build 34) on chromosome 4. The upper plot includes pairwise  $D'$  for 230 common SNPs [with MAF (minor allele frequency)  $> 5\%$ ] from the HapMap release 22 for the CEU population (24); the lower plot includes pairwise  $r^2$  values for the same set of SNPs. (B) Sex-averaged recombination rate (saRR) estimates (in centimorgans per megabase) in the same interval, on the basis of the HapMap dataset (26). (C) Location of nine known genes in this region. (D) Schematic view of the correlation between recombination rates and genotypes for SNPs in the interval from the genome-wide association study of 1887 males (blue dots) and 1702 females (red dots). Plotted is  $-\log_{10} P$ , where  $P$  is the adjusted  $P$  value, against the chromosomal location of the markers. All four panels use the same horizontal megabase scale indicated at the bottom of panel (D).

**Table 1.** Association of sequence variants with male and female recombination rate. The two alleles of each SNP, rs3796619, rs1670533, and rs2045065, are shown. The frequency of the first allele (e.g., allele T of

rs3796619) in the sample is given. Effect is estimated on the basis of an additive model. Displayed is the estimated effect of the first allele relative to the second allele.

Sample	SNP (alleles) (%)	Male recombination rate		Female recombination rate	
		Effect (cM) [95% CI]	P	Effect (cM) [95% CI]	P
Genome-wide					
1887M/1702F	rs3796619 (T/C) (32.9)	-67.9	$1.1 \times 10^{-14}$	67.6	$7.9 \times 10^{-6}$
	rs1670533 (C/T) (22.5)	-66.1	$1.8 \times 10^{-11}$	92.8	$4.1 \times 10^{-8}$
	rs2045065 (G/A) (22.5)	-66.2	$1.6 \times 10^{-11}$	92.2	$6.0 \times 10^{-8}$
Replication					
1248M/1663F	rs3796619 (T/C) (32.8)	-75.2	$3.5 \times 10^{-11}$	37.4	0.025
	rs1670533 (C/T) (23.0)	-74.5	$2.1 \times 10^{-9}$	82.9	$5.4 \times 10^{-6}$
	rs2045065 (G/A) (22.9)	-74.1	$2.8 \times 10^{-9}$	79.0	$1.6 \times 10^{-5}$
Combined					
3135M/3365F	rs3796619 (T/C) (32.9)	-70.7 [-84.3, -57.1]	$3.2 \times 10^{-24}$	54.3 [32.1, 76.5]	$1.6 \times 10^{-6}$
	rs1670533 (C/T) (22.7)	-69.4 [-84.6, -54.2]	$2.8 \times 10^{-19}$	88.2 [63.7, 112.7]	$1.9 \times 10^{-12}$
	rs2045065 (G/A) (22.6)	-69.3 [-84.5, -54.1]	$3.3 \times 10^{-19}$	86.1 [61.4, 110.8]	$8.6 \times 10^{-12}$



**Fig. 2.** (A and B) Difference in recombination rate between the three haplotypes of rs3796619 and rs1670533. Size of the circle is proportional to the frequency of the haplotype in the population (67.1% [C,T], 22.7% [T,C], 10.2% [T,T]). Each arrow indicates the comparison of two haplotypes and the difference of the haplotype the arrow is pointing to relative to the second haplotype.

effect on recombination rate for 22 of the 23 chromosomes, 13 of them significantly so. Unlike males, for females, a test of heterogeneity was not significant ( $P > 0.05$ ), but chromosome 21 stood out as one that gives an estimated effect in the opposite direction.

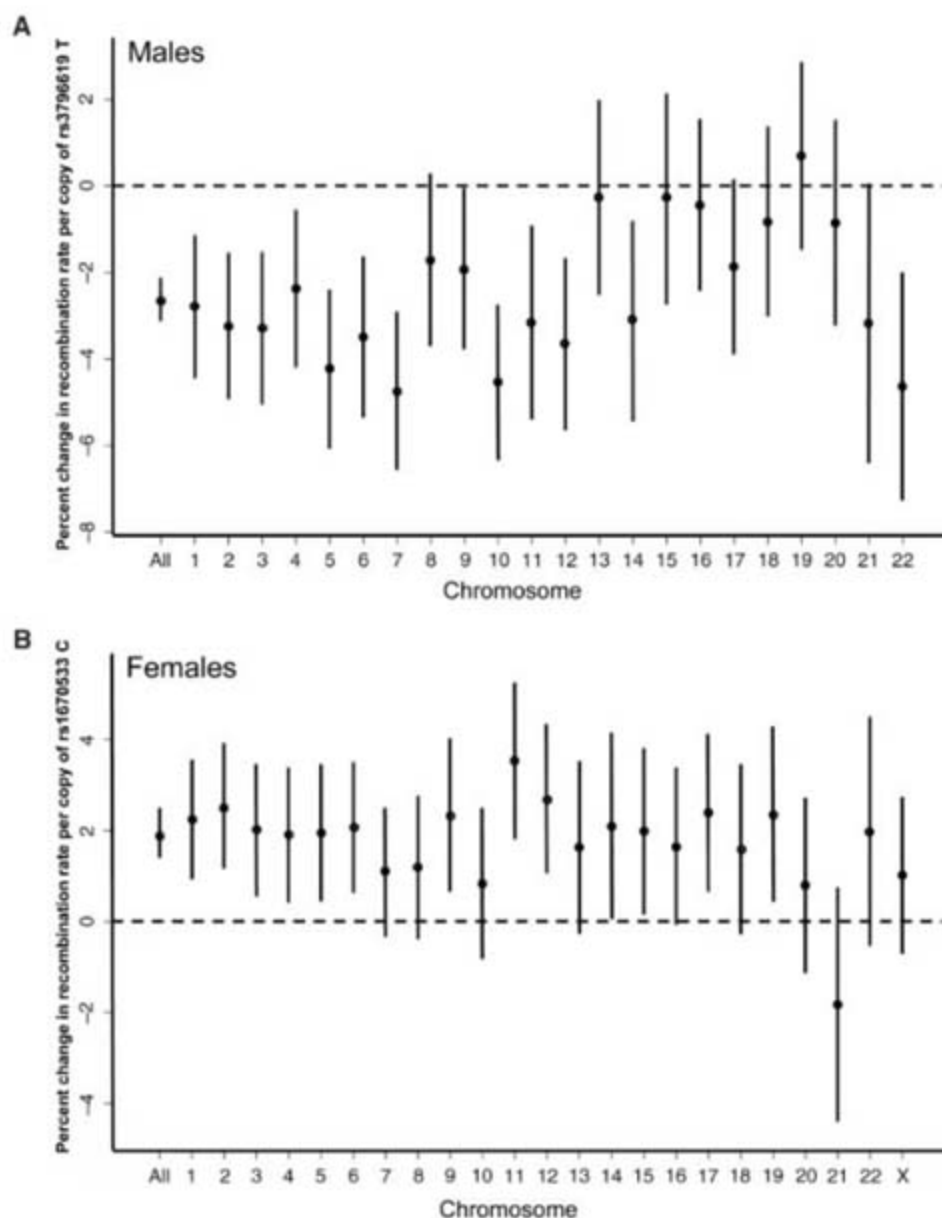
Relative to the population average, each copy of rs3796619 T was estimated to decrease male genome-wide recombination rate by 2.62%. For rs1670533 C and female genome-wide recombination rate, the estimated effect was an increase of 1.87%. From the regression results, rs3796619 T explained 3.5% of the variation in male recombination rate and rs1670533 C explained 1.7% of the variation in female recombination rate. These numbers, however, substantially underestimate the total parental effect explained by these variants. This is because, on the basis of the recombination counts of a few children only, the inherent recombination rate of an individual, which mathematically can be viewed as the average recombination counts in the children (if the person were to have an infinite number of children), was measured with substantial error. The latter has to be estimated and deducted from the total variation in the regression analyses in

order to properly evaluate the contribution of the identified variants. With a method that utilizes the different chromosomes as pseudo-replicates, we estimated that about 6.6% of the total variation in the paternal recombination count of a child could be attributed to a systematic effect associated with the father [see (19) for a description of the method and the decomposition of the total variation into various components]. A systematic maternal effect was estimated to account for about 11% of the variation in the maternal recombination count of a child. For the fathers and mothers in this study, who have, on average, about 2.75 genotyped children, the paternal effect accounted for ~16% of the variation in the estimated recombination rates of the fathers, and the maternal effect accounted for ~26% of the variation in the estimated recombination rates of the mothers. Hence, rs3796619 explains about 22% (3.5/16) of the paternal effect, and rs1670533 explains about 6.5% (1.7/26) of the maternal effect.

The three SNPs with the strongest association with recombination rate in males and females were located in an LD block spanning 200 kb. Two genes, *SPON2* and *RNF212*, are located

within this LD block (Fig. 1). The *SPON2* gene may act as an opsonin and pattern-recognition molecule for a range of pathogens through detection of carbohydrate structures and activation of macrophages. The *RNF212* gene has not been characterized, but homology searches indicate that the gene encodes a RING finger protein that may be a ubiquitin ligase. Gene ortholog predictions suggest homology between the mammalian *RNF212* gene and the *ZHP-3* gene (a homolog of the *ZIP3* gene in yeast), which is involved in meiotic recombination (21) (fig. S3). Meiotic recombination is likely to depend, in large part, on recombinatorial repair of programmed meiotic double-strand breaks (22). The synaptonemal complex (SC), a structure formed by close association of axes of homologous sister chromatid pairs, is essential for crossover formation and completion of meiosis (23). Three proteins involved in this complex, Zip2, Zip3, and Zip4 (from the yeast Zip genes), mediate protein-protein interactions (23). On the basis of gene knockout experiments, the *ZIP3* homolog in *Caenorhabditis elegans*, *ZHP-3* (K02B12.8) is essential for reciprocal recombination between homologous chromosomes and, thus, for chiasma formation (21). Similarities between distantly related organisms, such as *Saccharomyces cerevisiae* and *C. elegans*, suggest that the structure and function of the SC is conserved among distantly related species. This suggests that the *RNF212* gene may play a crucial role in recombination and assembly of the SC (21) in mammals, as do their putative orthologs in *C. elegans* and *S. cerevisiae*. Although the conservation between *RNF212* in humans and *ZHP-3* in *C. elegans* is limited (36 out of 118 amino acids in the RING finger domain) (21), our findings support the idea that the *RNF212* protein may be involved in recombination.

In the LD block containing *RNF212*, on the basis of the HapMap group composed of European Americans (CEU), data release 22 (24), there are a large number of SNPs highly corre-



**Fig. 3.** Effect on recombination rate for individual chromosomes and all chromosomes combined. Estimate and 95% confidence interval are displayed for percentage change. **(A)** Effect of allele T of rs3796619 in males. **(B)** Effect of allele C of rs1670533 in females.

lated with either rs3796619 or rs1670533 or with haplotypes formed by alleles of the two SNPs. Nine SNPs tag rs3796619 and 43 SNPs tag rs1670533 with a pairwise correlation coefficient  $r^2 > 0.9$ . One of the SNPs strongly correlated with rs3796619 is rs4045481 ( $r^2 = 0.96$ ), a silent (synonymous) SNP in the third exon of *RNF212* mRNA (BC050356). A deletion, rs33995490, in the coding region of the fourth exon of another *RNF212* mRNA (BC036250) is strongly correlated with allele C of rs1670533 ( $r^2 = 1.0$ ) (figs. S4 and S5). Thus, these variants may affect recombination rate, but functional work is required to prove causality. We sequenced the exons of the *RNF212* gene and did not find any other common coding variants in the RefSeq Gene variant (NM\_194439.1) that could account for the observed association with recombination rate (see table S2 for details).

A phylogenetic analysis of a 55-kb region containing rs3796619 and rs1670533 in the

HapMap data (24) revealed three well-differentiated clusters of haplotypes showing notable differences in frequency between the Yoruban Nigerians (YRI) and CEU and East Asians (CHB and JPT) (fig. S6). The [C,T] and [T,C] haplotypes that associate most strongly with recombination rate have a combined frequency of only 17% in the YRI sample, but reach a frequency of 91% and 98% in the CEU and East Asian samples, respectively. Several SNPs in this region show an unusual degree of divergence among the HapMap groups, on the basis of the rank percentile of their  $F_{ST}$  values (Wright's coefficient, a measure of variance in allele frequencies among populations) among all autosomal SNPs with the same overall frequency in the HapMap. Specifically, we identified eight SNPs whose  $F_{ST}$  values are in the top 0.5% for differences between the YRI and East Asian HapMap samples and also in the top 5% of differences between the YRI and CEU samples. Each of these SNPs differentiated a

subset of [T,T] haplotypes from the rest, perhaps indicating an episode of positive selection (or a severe founder effect) that increased the frequency of [C,T] and [T,C] haplotypes in the ancestors of European and East Asian populations.

Although an inversion on 17q21.31 has previously been associated with recombination rate in women, the variants described here are shown to associate with both male and female recombination rate. Notably, they are associated with opposite effects in the two sexes. It is possible that variants with such properties are unique to recombination rates. Biologically, the processes of male and female recombination are different (25), which may allow such variants to exist. These variants could serve a key function from an evolutionary perspective, as they allow the transfer of the recombination contribution from one sex to the other with minimal impact on the average recombination rate for the population as a whole.

#### References and Notes

- G. Coop, M. Przeworski, *Nat. Rev. Genet.* **8**, 23 (2007).
- B. S. Baker, A. T. Carpenter, M. S. Esposito, R. E. Esposito, L. Sandler, *Annu. Rev. Genet.* **10**, 53 (1976).
- T. Hassold, P. Hunt, *Nat. Rev. Genet.* **2**, 280 (2001).
- K. Inoue, J. R. Lupski, *Annu. Rev. Genomics Hum. Genet.* **3**, 199 (2002).
- W. Winckler *et al.*, *Science* **308**, 107 (2005).
- S. Myers, L. Bottolo, C. Freeman, G. McVean, P. Donnelly, *Science* **310**, 321 (2005).
- A. J. Jeffreys, R. Neumann, *Nat. Genet.* **31**, 267 (2002).
- A. J. Jeffreys, R. Neumann, *Hum. Mol. Genet.* **14**, 2277 (2005).
- A. J. Jeffreys, R. Neumann, M. Panayi, S. Myers, P. Donnelly, *Nat. Genet.* **37**, 601 (2005).
- M. Carrington, M. Cullen, *Trends Genet.* **20**, 196 (2004).
- I. Tiemann-Boege, P. Calabrese, D. M. Cochran, R. Sokol, N. Arnheim, *PLoS Genet.* **2**, e70 (2006).
- K. W. Broman, J. C. Murray, V. C. Sheffield, R. L. White, J. L. Weber, *Am. J. Hum. Genet.* **63**, 861 (1998).
- A. Kong *et al.*, *Nat. Genet.* **31**, 241 (2002).
- A. Lynn *et al.*, *Science* **296**, 2222 (2002).
- T. Hassold *et al.*, *Cytogenet. Genome Res.* **107**, 249 (2004).
- F. Sun *et al.*, *Hum. Mol. Genet.* **15**, 2376 (2006).
- A. Kong *et al.*, *Nat. Genet.* **36**, 1203 (2004).
- H. Stefansson *et al.*, *Nat. Genet.* **37**, 129 (2005).
- Materials and methods are available as supporting material on Science Online.
- B. Devlin, K. Roeder, *Biometrics* **55**, 997 (1999).
- V. Jantsch *et al.*, *Mol. Cell. Biol.* **24**, 7998 (2004).
- Y. Costa, H. J. Cooke, *Chromosome Res.* **15**, 579 (2007).
- A. Lynn, R. Soucek, G. V. Borner, *Chromosome Res.* **15**, 591 (2007).
- The International HapMap Consortium, *Nature* **437**, 1299 (2005).
- P. A. Hunt, T. J. Hassold, *Science* **296**, 2181 (2002).
- G. A. McVean *et al.*, *Science* **304**, 581 (2004).
- Sequences were deposited in GenBank database accession numbers EU316220–EU325532. SNPs identified through sequencing are listed in table S2 with accession numbers.

#### Supporting Online Material

[www.sciencemag.org/cgi/content/full/1152422/DC1](http://www.sciencemag.org/cgi/content/full/1152422/DC1)  
Materials and Methods  
Figs. S1 to S6  
Tables S1 to S3  
References

31 October 2007; accepted 2 January 2008  
Published 31 January 2008;  
10.1126/science.1152422  
Include this information when citing this paper.

# Hepatic Glucose Sensing via the CREB Coactivator CRTC2

Renaud Dentin,<sup>1</sup> Susan Hedrick,<sup>1</sup> Jianxin Xie,<sup>2</sup> John Yates III,<sup>3</sup> Marc Montminy<sup>1\*</sup>

Chronic hyperglycemia contributes to the development of diabetes-associated complications. Increases in the concentration of circulating glucose activate the hexosamine biosynthetic pathway (HBP) and promote the O-glycosylation of proteins by O-glycosyl transferase (OGT). We show that OGT triggered hepatic gluconeogenesis through the O-glycosylation of the transducer of regulated cyclic adenosine monophosphate response element-binding protein (CREB) 2 (TORC2 or CRTC2). CRTC2 was O-glycosylated at sites that normally sequester CRTC2 in the cytoplasm through a phosphorylation-dependent mechanism. Decreasing amounts of O-glycosylated CRTC2 by expression of the deglycosylating enzyme O-GlcNAcase blocked effects of glucose on gluconeogenesis, demonstrating the importance of the HBP in the development of glucose intolerance.

In fasted animals, increases in the concentration of circulating pancreatic glucagon trigger the gluconeogenic program in part by stimulating the dephosphorylation of CRTC2 at Ser 171 (1, 2). CRTC2 is sequestered in the cytoplasm through an association with 14-3-3 proteins in animals fed ad libitum but translocates to the nucleus after its dephosphorylation, where it mediates transcriptional activation by binding to CREB.

In response to refeeding, increases in circulating concentrations of pancreatic insulin inhibit hepatic glucose production in part through the ubiquitin-dependent degradation of CRTC2 (3). The gluconeogenic program is constitutively induced in diabetes because of insulin resistance and chronic increases in circulating concentrations of glucose, which appear to increase gluconeogenic gene expression independently (4–6). Because a conserved cyclic adenosine 3'-5' monophosphate (cAMP) response element (CRE) on the glucose-6-phosphatase (G6Pase) promoter is required for the transcription of that gene in response to glucose (7), we explored the role of CRTC2 in this process.

Exposure of primary mouse hepatocytes to various concentrations of glucose (Glu) or glucosamine (GlcN), an intermediate in the HBP, stimulated gluconeogenic gene expression (G6Pase and PEPCK) and increased glucose output from primary mouse hepatocytes; these effects were augmented by the adenylyl cyclase activator forskolin (FSK) (Fig. 1, A and B, and fig. S1). CRTC2 appeared critical in this regard because Glu and GlcN did not promote the accumulation of G6Pase or PEPCK mRNAs in cells expressing CRTC2 RNA interference (RNAi). Exposure to Glu or GlcN also increased wild-type G6Pase promoter activity cooperatively with FSK in hepatocytes but had no effect on a CRE-mutant G6Pase reporter

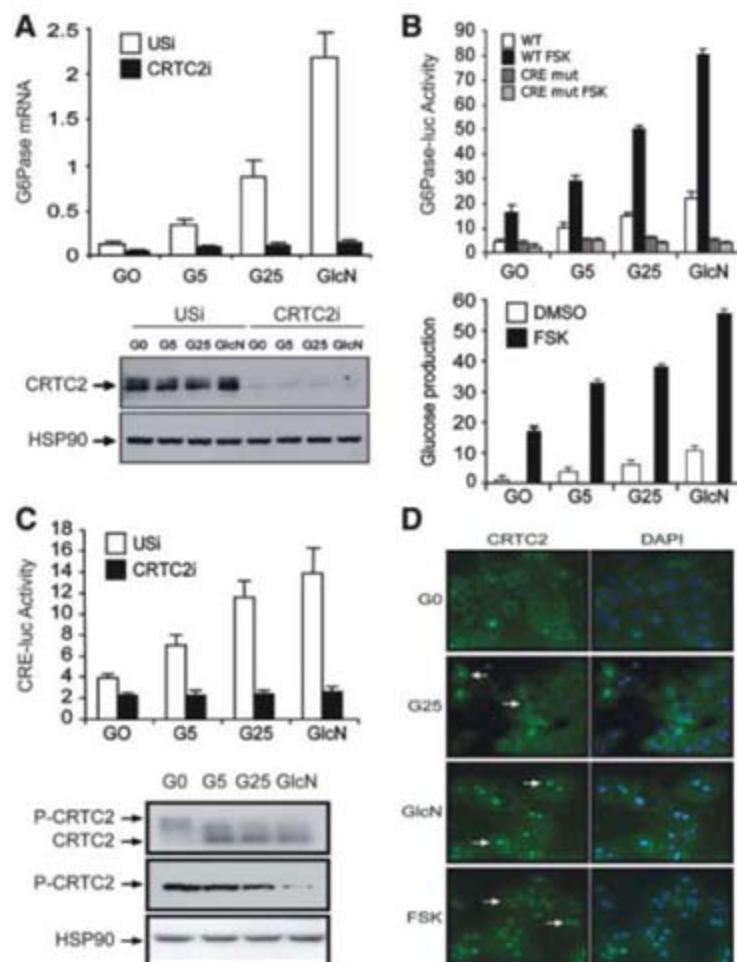
(Fig. 1B). Further demonstrating the ability of the CREB:CRTC2 pathway to mediate glucose-responsive transcription, treatment of cells with Glu or GlcN increased the activity of a CRE-luciferase (CRE-luc) reporter containing only CREB binding sites in control cells but not in CRTC2-deficient cells (Fig. 1C). In line with these changes, exposure of cells to Glu or GlcN also promoted the dephosphorylation and nu-

clear accumulation of CRTC2, as detected by immunoblotting and immunofluorescence analyses (Fig. 1, C and D, and fig. S1).

Increases in the concentration of extracellular glucose modulate gene expression in part by triggering the O-glycosylation of certain transcriptional activators (8, 9). Thus, we investigated whether CRTC2 is O-glycosylated in hepatocytes. Amounts of O-glycosylated CRTC2 (OG-CRTC2) increased after treatment with Glu or related analogs (mannose, fructose, or galactose), as detected by Western blotting of CRTC2 immunoprecipitated with antiserum that recognizes O-glycosylated proteins (10) (Fig. 2A and fig. S2). Glucose analogs and GlcN also stimulated CRE-luc activity in primary hepatocytes, but nonmetabolizable sugars (2-deoxyglucose and L-glucose) and glycolytic intermediates downstream of the HBP did not.

To determine whether the HBP is required for glucose-stimulated activation of CRTC2, we used inhibitors of L-glutamine fructose-6-phosphate amido transferase (GFAT), the rate-limiting enzyme in this pathway (11). Compared with hepatocytes exposed to high glucose alone, cells incubated with the GFAT inhibitors 6-diazo-5-oxo-L-norleucine (DON) or azaserine (AZA) had

**Fig. 1.** Activation of CRTC2 in cells exposed to high concentrations of glucose. (A) (Top) Effect of glucose (0, 5, 25 mM: G0, G5, G25) and glucosamine (10mM GlcN) exposure (6 hours) on amounts of G6Pase mRNA in primary mouse hepatocytes infected with adenovirally encoded unspecific (USi) or adenovirally encoded CRTC2 RNAi (CRTC2i). (Bottom) Western blot showing effect of CRTC2 knock-down on amounts of CRTC2 protein relative to control cells. (For this and subsequent figures, the *P* value was determined using a two-tailed unpaired Student's *t* test. *P* < 0.05 for USi compared with CRTC2i cells; *n* = 3). (B) Effect of Glu and GlcN with or without FSK (10  $\mu$ M, 6 hours) on wild-type and CRE-mutant G6Pase-luciferase reporters (top) and on glucose output (bottom) from primary mouse hepatocytes. Cells were preincubated in glucose for 6 hours, washed, and then assayed for glucose release into glucose-free medium over a 1-hour period. (C) (Top) Effect of Glu and GlcN on CRE-luciferase reporter activity in USi- and CRTC2i-expressing mouse hepatocytes, as in (A). (*P* < 0.05 relative to control cells; *n* = 3). (Bottom) Western blot of phospho (Ser<sup>171</sup>) and total CRTC2 in hepatocytes exposed to Glu or GlcN. (D) Immunofluorescence analysis of CRTC2 localization in hepatocytes exposed to Glu, GlcN, or FSK for 6 hours and stained for endogenous CRTC2. DAPI staining shown to visualize nuclei. White arrows point to representative cells with nuclear CRTC2 staining.



<sup>1</sup>The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA. <sup>2</sup>Cell Signaling Technology, 3 Trask Lane, Danvers, MA 01923, USA. <sup>3</sup>The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA.

\*To whom correspondence should be addressed. E-mail: montminy@salk.edu

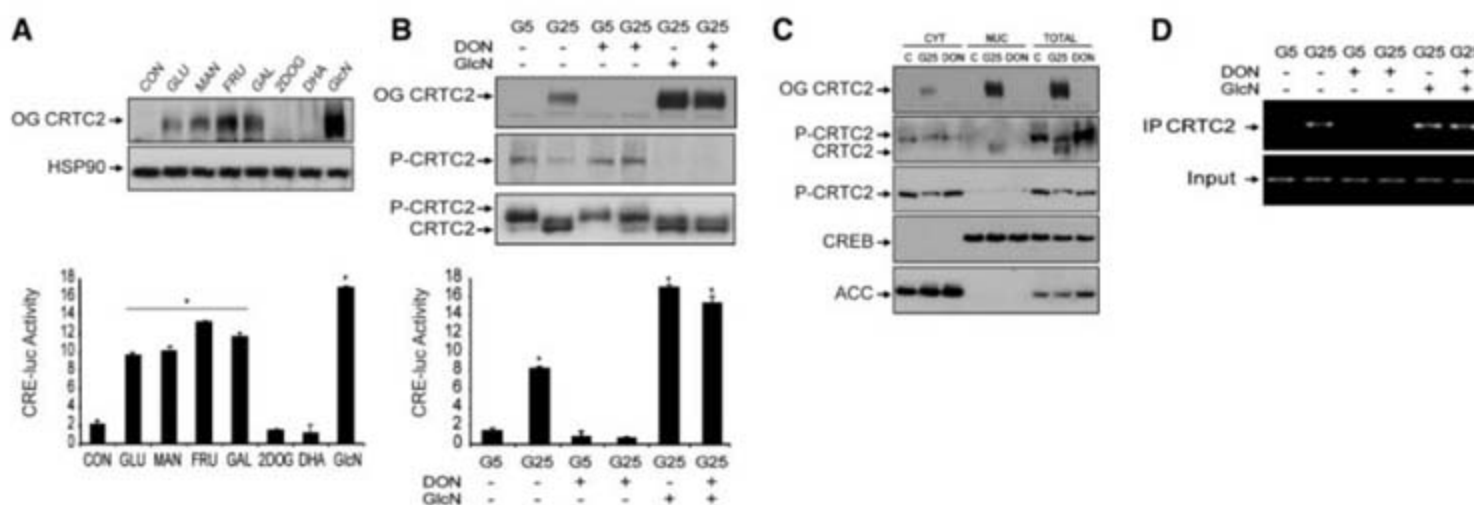
lower amounts of OG-CRTC2 (Fig. 2B and fig. S3). DON and AZA also inhibited CRTC2 nuclear translocation and CRE-luc activation by glucose, supporting the idea that the HBP mediates effects of glucose on CRTC2 activation in hepatocytes.

O-glycosylation and phosphorylation often exert reciprocal effects on protein activity (12). Indeed, phosphorylated CRTC2 (P-CRTC2), detected by immunoblotting, was confined to the cytoplasm, whereas OG-CRTC2 was detected primarily in nuclear fractions of glucose-stimulated hepatocytes (Fig. 2C). In chromatin immunoprecipitation (ChIP) assays, amounts of OG-CRTC2 associated with the G6Pase promoter increased after exposure of cells to high concentrations of

glucose (25 mM) (Fig. 2D and fig. S4). Blocking CRTC2 O-glycosylation with DON disrupted the effects of glucose on CRTC2 recruitment; and co-incubation of DON-treated cells with GlcN rescued both CRTC2 O-glycosylation and occupancy of the G6Pase promoter.

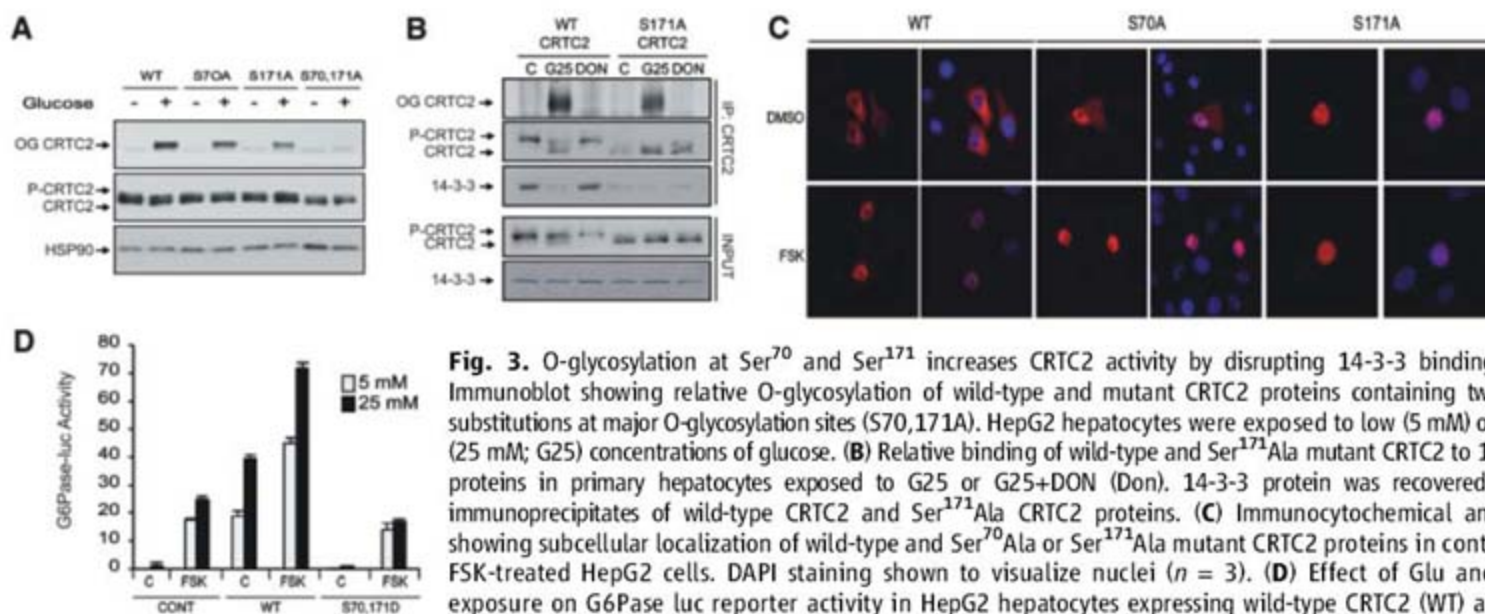
In mass spectrometry studies to characterize O-glycosylation sites in CRTC2, we identified six residues, two of which (Ser<sup>70</sup> and Ser<sup>171</sup>) showed greater amounts of O-glycosylation by mutational analysis (Fig. 3A and fig. S5). A mutant CRTC2 protein containing Ala substitutions at both Ser<sup>70</sup> and Ser<sup>171</sup> was not detectably O-glycosylated after exposure of hepatocytes to high glucose concentrations. In resting cells, CRTC2 is

sequestered in the cytoplasm by 14-3-3 proteins through phosphorylation at Ser<sup>70</sup> and Ser<sup>171</sup> by members of the AMP-activated protein kinase (AMPK) family of Ser-Thr kinases (13, 2). Consistent with the proposed role of O-glycosylation in blocking protein phosphorylation, alanine mutations at Ser<sup>70</sup> and Ser<sup>171</sup> disrupted the CRTC2:14-3-3 interaction and promoted localization of these mutant CRTC2 proteins to the nucleus (Fig. 3, B and C, and fig. S6). Phosphorylation-defective Ser<sup>70</sup>Ala and Ser<sup>171</sup>Ala mutant CRTC2 proteins were also more active than wild-type CRTC2 in stimulating CRE-luc activity under basal conditions (fig. S5). Conversely, substitution of Ser<sup>70</sup> and Ser<sup>171</sup> with aspartate, to mimic constitutive



**Fig. 2.** Glucose stimulated O-glycosylation, nuclear translocation, and recruitment of CRTC2 to gluconeogenic genes. (A) Effect of Glu (25 mM) and related analogs (mannose, fructose, or galactose) or the nonmetabolizable analog 2-Deoxyglucose (DOG), the HBP intermediate GlcN, or the downstream glycolytic intermediate dihydroxyacetone phosphate (DHA) on amounts of O-glycosylated CRTC2 (OG-CRTC2) (top) and on CRE-luc activity (bottom) in primary mouse hepatocytes. (\*,  $P < 0.05$  compared to control;  $n = 3$ ). (B) Effect of GFAT inhibitor DON on amounts of OG-CRTC2 (top) and on CRE-luc activity (bottom) in hepatocytes exposed to 5 mM glucose (G5), 25 mM glucose (G25), or GlcN

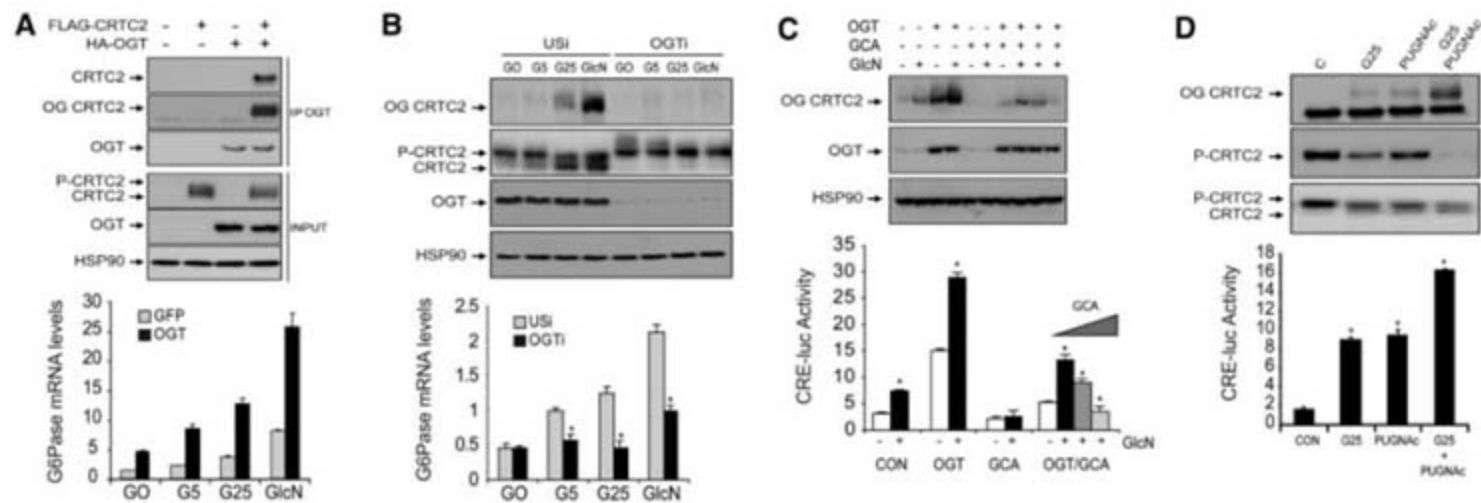
for 6 hours. Amounts of phospho (Ser<sup>171</sup>) CRTC2 (P-CRTC2) and total CRTC2 are shown. (\*,  $P < 0.05$  compared with control;  $n = 3$ ). (C) Western blot showing amounts of OG-CRTC2 and P-CRTC2 in nuclear and cytoplasmic fractions of primary hepatocytes exposed to G25, G25 and DON (DON), or G5 (C). Control nuclear (CREB) and cytoplasmic (ACC) proteins shown. (D) ChIP assay showing effect of low (5 mM) or high (25 mM) concentrations of glucose on recruitment of CRTC2 to the G6Pase promoter in hepatocytes. Effects of GFAT inhibitor (DON) and glucosamine (GlcN) are indicated ( $n = 2$ ).



**Fig. 3.** O-glycosylation at Ser<sup>70</sup> and Ser<sup>171</sup> increases CRTC2 activity by disrupting 14-3-3 binding. (A) Immunoblot showing relative O-glycosylation of wild-type and mutant CRTC2 proteins containing two Ala substitutions at major O-glycosylation sites (S70,171A). HepG2 hepatocytes were exposed to low (5 mM) or high (25 mM; G25) concentrations of glucose. (B) Relative binding of wild-type and Ser<sup>171</sup>Ala mutant CRTC2 to 14-3-3 proteins in primary hepatocytes exposed to G25 or G25+DON (Don). 14-3-3 protein was recovered from immunoprecipitates of wild-type CRTC2 and Ser<sup>171</sup>Ala CRTC2 proteins. (C) Immunocytochemical analysis showing subcellular localization of wild-type and Ser<sup>70</sup>Ala or Ser<sup>171</sup>Ala mutant CRTC2 proteins in control or FSK-treated HepG2 cells. DAPI staining shown to visualize nuclei ( $n = 3$ ). (D) Effect of Glu and FSK exposure on G6Pase luc reporter activity in HepG2 hepatocytes expressing wild-type CRTC2 (WT) and O-glycosylation defective (S70A, S171A; S70,171D) CRTC2 constructs, or empty vector control (CON).

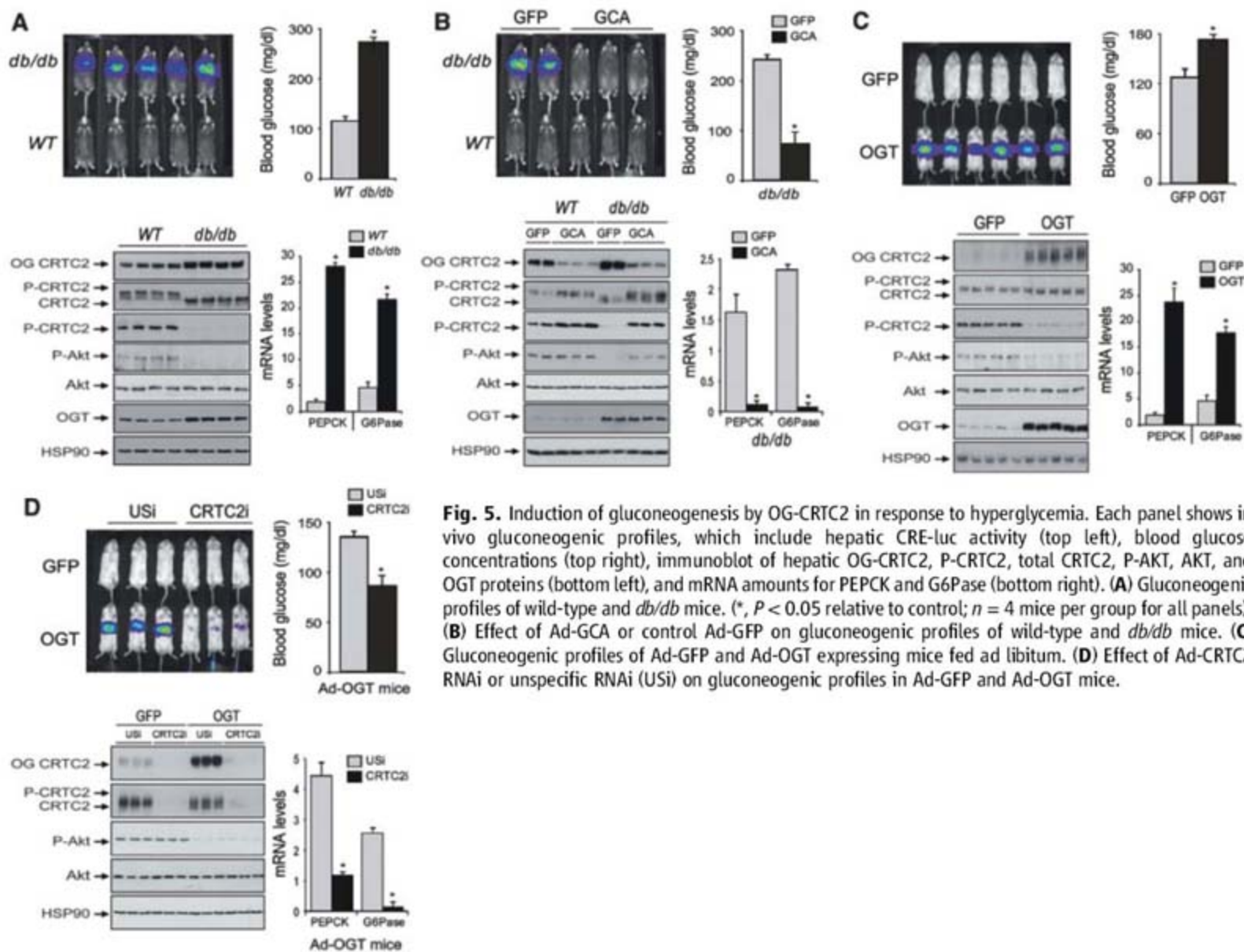
Comparable expression of wild-type and mutant CRTC2 proteins was verified by Western blot. (\*,  $P < 0.05$  compared with 5 mM glucose;  $n = 3$ ).





**Fig. 4.** Modulation of O-glycosylation and transcriptional activity of CRTC2 by OGT and GCA. **(A)** (Top) Coimmunoprecipitation assay of HEK293T cells using epitope-tagged CRTC2 and OGT proteins. Amounts of CRTC2 recovered from IPs of OGT shown. O-glycosylation of CRTC2 (OG-CRTC2) by OGT was detected by immunoblotting. (Bottom) Effect of Ad-OGT or control Ad-GFP on amounts of G6Pase mRNA in primary hepatocytes exposed to various concentrations (0, 5, or 25 mM) of glucose (Glu) or glucosamine (GlcN) (10 mM) for 6 hours. ( $P < 0.05$  for OGT compared with GFP-expressing cells;  $n = 3$ ). **(B)** Effect of OGT small interfering RNA or unspecific control (USi) on amounts of OGT or

amounts of OG-CRTC2 (top) and on gluconeogenic gene expression (G6Pase) (bottom) in hepatocytes exposed to various concentrations of Glu or GlcN. ( $*$ ,  $P < 0.05$  relative to control;  $n = 3$ ). **(C)** Effect of Ad-O-GlcNAcase (GCA), either alone or in combination with Ad-OGT on CRTC2 O-glycosylation (top) and CRE-luc activity (bottom) in hepatocytes exposed to GlcN as indicated for 6 hours. Wedge shape indicates increasing amounts of Ad-GCA. ( $*$ ,  $P < 0.05$  relative to control;  $n = 3$ ). **(D)** Effect of GCA inhibitor PUGNAc on amounts of OG-CRTC2 (top) and CRE-luc activity (bottom) in cells exposed to 5 mM or 25 mM Glu ( $*$ ,  $P < 0.05$  relative to control;  $n = 3$ ).



**Fig. 5.** Induction of gluconeogenesis by OG-CRTC2 in response to hyperglycemia. Each panel shows in vivo gluconeogenic profiles, which include hepatic CRE-luc activity (top left), blood glucose concentrations (top right), immunoblot of hepatic OG-CRTC2, P-CRTC2, total CRTC2, P-AKT, AKT, and OGT proteins (bottom left), and mRNA amounts for PEPCK and G6Pase (bottom right). **(A)** Gluconeogenic profiles of wild-type and *db/db* mice. ( $*$ ,  $P < 0.05$  relative to control;  $n = 4$  mice per group for all panels). **(B)** Effect of Ad-GCA or control Ad-GFP on gluconeogenic profiles of wild-type and *db/db* mice. **(C)** Gluconeogenic profiles of Ad-GFP and Ad-OGT expressing mice fed ad libitum. **(D)** Effect of Ad-CRTC2 RNAi or unspecific RNAi (USi) on gluconeogenic profiles in Ad-GFP and Ad-OGT mice.

phosphorylation at these sites, disrupted effects of glucose on CRTC2 activity in hepatocytes, confirming the importance of these sites for transcriptional regulation in response to glucose (Fig. 3D).

O-glycosyl transferase (OGT) catalyzes the O-glycosylation of cellular proteins in response to activation of the HBP (9). In proteomic studies to identify CRTC2-associated proteins, we recovered OGT from IPs of CRTC2 (fig. S7). We confirmed the CRTC2:OGT interaction in IP studies using epitope-tagged OGT and CRTC2 constructs (Fig. 4A). Overexpression of OGT increased amounts of OG-CRTC2 and stimulated CRE-luc activity along with gluconeogenic gene expression (G6Pase) in hepatocytes (Fig. 4A and fig. S8). Conversely, RNAi-mediated knockdown of OGT blocked the effects of Glu or GlcN on CRTC2 glycosylation and on gluconeogenic gene expression (Fig. 4B).

Protein O-glycosylation by OGT is rapidly reversible in vivo through opposing effects of the deglycosylating enzyme O-GlcNAcase (GCA) (12). Expression of adenovirally encoded GCA (Ad-GCA) in hepatocytes reduced amounts of OG-CRTC2 and disrupted CRE-luc activity in response to GlcN and to Ad-OGT (Fig. 4C). Conversely, treating cells with GCA inhibitors O-(2-acetamido-2-deoxy-D-glucopyranosylidene) amino N-phenyl carbamate (PUGNAc) or streptozotocin (STZ) (14) increased amounts of OG-CRTC2 and stimulated CRE-luc activity (Fig. 4D and fig. S9). These results support the notion that OGT and GCA exert counter-regulatory effects on CRTC2 O-glycosylation and activation in hepatocytes.

We evaluated whether chronic increases in circulating glucose concentrations are suffi-

cient to trigger CRTC2 O-glycosylation and gluconeogenic gene expression in vivo using insulin resistant *db/db* diabetic mice and mice fed a high-fat diet (HFD). *Db/db* and HFD mice had higher gluconeogenic profiles, which include hepatic CRE-luc activity, gluconeogenic gene expression, circulating glucose concentrations, and amounts of hepatic OG-TORC2, than did control animals (Fig. 5A and figs. S10 and S11). Disrupting CRTC2 O-glycosylation in HFD and *db/db* animals through expression of hepatic Ad-GCA lowered the gluconeogenic profile (Fig. 5B and fig. S12). As a result, HFD and *db/db* mice expressing Ad-GCA showed increased glucose tolerance and insulin sensitivity. Although GCA could improve glucose homeostasis by deglycosylating components of the insulin signaling pathway, Ad-GCA expression in liver down-regulated the gluconeogenic profile comparably to Ad-CRTC2i in streptozotocin-diabetic mice, in which hepatic insulin signaling is absent as a result of the destruction of insulin-producing pancreatic beta cells (fig. S13). Conversely, increasing OG-CRTC2 amounts through expression of Ad-OGT in liver enhanced gluconeogenic profiles in wild-type mice (Fig. 5C and fig. S14). We tested whether CRTC2 was required for Ad-OGT-mediated induction of the gluconeogenic program in RNAi knockdown studies. Relative to control Ad-OGT animals expressing unspecific RNAi (Ad-USi), Ad-OGT mice coinjected with Ad-CRTC2i had lower gluconeogenic profiles (Fig. 5D and fig. S14).

Chronic hyperglycemia is thought to contribute to the development of diabetes-associated complications in part by activating the HBP and

increasing protein O-glycosylation at regulatory phosphorylation sites (12, 15). Reducing the O-glycosylation of CRTC2 and other metabolic regulators may improve glucose homeostasis and reduce long-term complications associated with this disease.

#### References and Notes

1. S. H. Koo *et al.*, *Nature* **437**, 1109 (2005).
2. R. A. Sreter *et al.*, *Cell* **119**, 61 (2004).
3. R. Dentin *et al.*, *Nature* **449**, 366 (2007).
4. D. Argaud, T. L. Kirby, C. B. Newgard, A. J. Lange, *J. Biol. Chem.* **272**, 12854 (1997).
5. J. K. Kim *et al.*, *J. Clin. Invest.* **108**, 153 (2001).
6. J. Shao, L. Qiao, R. C. Janssen, M. Pagliassotti, J. E. Friedman, *Diabetes* **54**, 976 (2005).
7. K. B. Pedersen *et al.*, *Am. J. Physiol. Endocrinol. Metab.* **292**, E788 (2007).
8. X. Cheng, R. N. Cole, J. Zaia, G. W. Hart, *Biochemistry* **39**, 11609 (2000).
9. J. E. Kudlow, *J. Cell. Biochem.* **98**, 1062 (2006).
10. S. A. Whelan, G. W. Hart, *Methods Enzymol.* **415**, 113 (2006).
11. M. G. Buse, *Am. J. Physiol. Endocrinol. Metab.* **290**, E1 (2006).
12. K. Kamemura, G. W. Hart, *Prog. Nucleic Acid Res. Mol. Biol.* **73**, 107 (2003).
13. A. K. Al-Hakim *et al.*, *J. Cell Sci.* **118**, 5661 (2005).
14. R. J. Konrad, K. M. Janowski, J. E. Kudlow, *Biochem. Biophys. Res. Commun.* **267**, 26 (2000).
15. S. Marshall, *Sci. STKE* **2006**, re7 (2006).
16. This work was supported by NIH grant RO1 GM037828, by the Clayton Medical Research Foundation, Inc., and by the Kieckhefer Foundation.

#### Supporting Online Material

[www.sciencemag.org/cgi/content/full/319/5868/1402/DC1](http://www.sciencemag.org/cgi/content/full/319/5868/1402/DC1)

Materials and Methods

Figs. S1 to S14

References

4 October 2007; accepted 25 January 2008

10.1126/science.1151363

## Coiled-Coil Irregularities and Instabilities in Group A *Streptococcus* M1 Are Required for Virulence

Case McNamara,<sup>1\*</sup> Annelies S. Zinkernagel,<sup>2</sup> Pauline Macheboeuf,<sup>1</sup> Madeleine W. Cunningham,<sup>3</sup> Victor Nizet,<sup>2,4</sup> Partho Ghosh<sup>1,5†</sup>

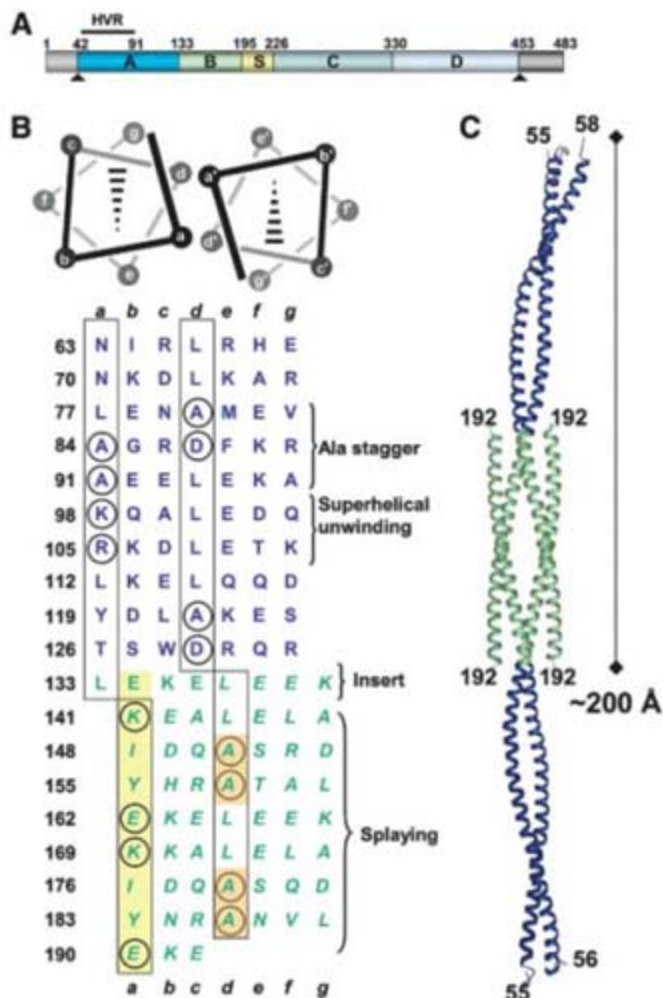
Antigenically variable M proteins are major virulence factors and immunogens of the human pathogen group A *Streptococcus* (GAS). Here, we report the ~3 angstrom resolution structure of a GAS M1 fragment containing the regions responsible for eliciting type-specific, protective immunity and for binding fibrinogen, which promotes M1 proinflammatory and antiphagocytic functions. The structure revealed substantial irregularities and instabilities throughout the coiled coil of the M1 fragment. Similar structural irregularities occur in myosin and tropomyosin, explaining the patterns of cross-reactivity seen in autoimmune sequelae of GAS infection. Sequence idealization of a large segment of the M1 coiled coil enhanced stability but diminished fibrinogen binding, proinflammatory effects, and antibody cross-reactivity, whereas it left protective immunogenicity undiminished. Idealized M proteins appear to have promise as vaccine immunogens.

**M** proteins are major virulence factors of group A *Streptococcus* (GAS), a bacterial pathogen responsible for mild-to-life-threatening diseases against which no vac-

cines currently exist (1). Fibrils of ~500 Å-long M protein form a dense, covalently attached coat on the streptococcal surface (2, 3). Host proteins, such as fibrinogen (4), bind specifically

to M proteins and block deposition of opsonic antibodies and complement, preventing phagocytic elimination of GAS by neutrophils (1, 5). A clone expressing the M1 antigenic variant of M protein emerged nearly three decades ago and has persisted as the leading cause of severe invasive GAS infection (6). Intact M1 and M1 fragments released by neutrophil proteases are sufficient to evoke pulmonary hemorrhage, inflammation, and tissue destruction that is characteristic of severe infection (7). These effects depend on M1 binding to fibrinogen, which triggers release of heparin binding protein (HBP), a mediator of vascular leakage, from neutrophils (7).

M proteins are also prominently associated with autoimmune sequelae of GAS infection, such as rheumatic fever, which is problematic for vaccine development (8) and remains a serious threat in the developing world. In rheumatic fever patients, potentially immunogenic M proteins elicit cross-reactive antibodies and T cell receptors directed against host  $\alpha$ -helical coiled-coil proteins, such as myosin and tropomyosin (1). Cross-reactivity is probably attributable to molecular mimicry, as M proteins appear to form coiled coils as well (2, 3, 9, 10). As with



**Fig. 1.** (A) Mature M1 results from cleavage (arrowheads) of the N-terminal signal sequence and the C-terminal Leu-Pro-X-Thr-Gly motif (where X is any amino acid) and covalent attachment of the C terminus to the cell wall. Boundaries of the A region, B repeats, S region, C repeats, and D region are indicated. (B) (Top) *a-d'* and *d-d'* (prime refers to the opposing helix) packing in parallel dimeric coiled coils. Broken wedges indicate helices pointing the N to C termini into the page. (Bottom) Heptad register indicated above and below the sequence (*a* and *d* position residues boxed) of the M1 A region (blue) and B repeats (green). The circled residues are destabilizing to coiled coils, with relative instabilities  $\Delta\Delta G_{\text{Ala}} \leq 0$  (11). Italicized residues form antiparallel coiled coils in the crystal. Residues highlighted in yellow and orange were substituted with Val and Leu, respectively, to create M1\* and M1AB\*. (C) Tail-to-tail packing of the two M1AB dimers in the asymmetric unit of the crystal (blue, A regions; green, B repeats).

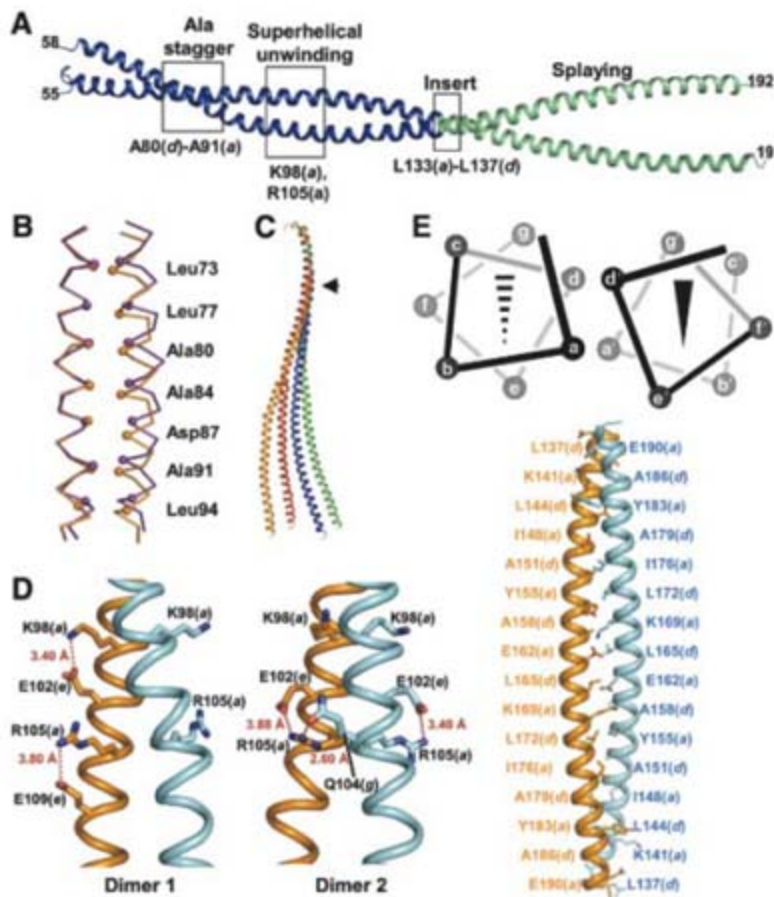
myosin and tropomyosin, M proteins contain coiled-coil destabilizing sequences (11–13)—that is, insertions within heptads and charged residues and Ala residues at *a* and *d* heptad positions (Fig. 1).

To understand the effects of such unusual sequence features in M proteins, we crystallized a fragment of M1 (called M1<sup>AB</sup>, residues 42 to 194) (14). The M1<sup>AB</sup> fragment contains the A region, whose first 50 residues, known as the hypervariable region (HVR), elicit type-specific, protective antibodies (5) and are part of a promising multivalent vaccine in clinical trials (15). The fragment also contains the B repeats, which are implicated in fibrinogen binding (4) and were sufficient to bind fibrinogen fragment D (FgD) (16) (fig. S1). M1<sup>AB</sup> is similar to a proinflammatory fragment generated by neutrophil proteases (7).

The 3.04 Å resolution structure of M1<sup>AB</sup> revealed that, whereas most of the A region formed a dimeric, parallel coiled coil, the B repeats had splayed apart and intertwined with the B repeats of adjacent M1<sup>AB</sup> molecules via antiparallel coiled coils (Fig. 1C, fig. S2, and table S1). The antiparallel association was probably an artifact of crystallization but is suggestive of instabilities in the B repeats.

Except for two short stretches of ideal parallel coiled coil (residues 63 to 79 and 106 to 119), the structure of M1<sup>AB</sup> was irregular throughout its ~200 Å length (Fig. 2A). The first of four major irregularities was an Ala stagger in the HVR. Poor packing of three Ala residues clustered at *a* and *d* positions led to local deformities; that is, a tightening of the coiled-coil radius from 5.0 to 4.25 Å, a ~2.5 Å asymmetric staggering of opposing helices, and a flexible hinge (Fig. 2, B and C, and fig. S3). Similar staggers and bends occur in tropomyosin (17, 18) and cardiac myosin (19) and are suggested to provide flexibility for function.

The second form of irregularity was superhelical unwinding due to Lys<sup>98</sup> and Arg<sup>105</sup> at successive *a* positions. These residues faced away from the coiled-coil core and contacted solvent-exposed residues (Fig. 2D), resulting in a loosening of the coiled-coil pitch from 150 Å to ~200 to 225 Å and an expansion of the coiled-coil radius to 5.4 Å (fig. S3). Unwinding resulting from Lys and Arg residues at *a* positions has been implicated in myosin function (19, 20) and also occurs in tropomyosin (17).



**Fig. 2.** (A) Structure of M1<sup>AB</sup> (blue, A region; green, B repeats) with boxed regions and labeling indicating irregularities. (B) Ala stagger shown by superposition of C $\alpha$  traces of M1<sup>AB</sup> residues 70 to 97 (orange) with the ideal coiled coil of GCN4 (purple). (C) Conformation of individual helices from the two M1<sup>AB</sup> dimers in the asymmetric unit, superimposed on main-chain atoms of residues 60 to 77. The position of the Ala stagger is indicated by the arrowhead. (D) Conformation of Lys<sup>98</sup> and Arg<sup>105</sup> in the two M1<sup>AB</sup> dimers in the asymmetric unit, with heptad positions of residues indicated in parentheses and polar contacts in red dashed lines (with distances shown). (E) (Top) Schematic of *a-d'* and *d-a'* packing in antiparallel dimeric coiled

coils. The broken wedge indicates the helix pointing the N to C termini into the page, and the solid wedge denotes out of the page. (Bottom) Antiparallel coiled coil of B repeats, with side chains of *a* and *d* position residues depicted and labeled.

<sup>1</sup>Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92093, USA. <sup>2</sup>Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093, USA. <sup>3</sup>University of Oklahoma Health Sciences Center, Biomedical Research Center, 975 North East 10th Street, Oklahoma City, OK 73104, USA. <sup>4</sup>School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093, USA. <sup>5</sup>Section of Molecular Biology, University of California, San Diego, La Jolla, CA 92093, USA.

\*Present address: Genomics Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Drive, San Diego, CA 92121, USA.

†To whom correspondence should be addressed. E-mail: pghosh@ucsd.edu

The third irregularity was attributable to an extra residue in the first heptad of the B repeats (Fig. 1B). The destabilizing effect of eight residues in a heptad (13) was accommodated by a +1 frameshift in the heptad register, precluding continuation of the parallel coiled coil. The fourth irregularity followed with the splaying apart of the B repeats and the formation of antiparallel coiled coils. The antiparallel orientation, with its *a-d'* (prime refers to the opposing helix) core packing (Fig. 2E), was probably preferable to the parallel orientation with its *a-a'* charge-charge clashes and *d-d'* Ala-Ala packing. Splaying at the ends of myosin (19) and tropomyosin (21, 22) coiled coils also occurs and is implicated in function.

Consistent with the prevalence of structural irregularities in M1<sup>AB</sup>, the circular dichroism (CD) spectrum of this fragment at 37°C showed a marked loss in  $\alpha$ -helical content and a 222:208 nm ratio < 1 (Fig. 3A). Because this ratio is  $\geq 1$  for coiled coils and  $\leq 0.86$  for isolated helices (23), these data suggested that M1<sup>AB</sup> exchanges between monomer and dimer states. This conclusion was supported by static light-scattering measurements, which provided evidence for the coexistence of M1<sup>AB</sup> monomers and dimers (fig. S4).

Intact M1 (residues 42 to 453) showed a comparable loss of  $\alpha$ -helical content at 37°C (Fig. 3B) (10). To determine whether monomer/dimer exchange also occurred in intact M1, we incubated His<sub>6</sub>-tagged M1 dimers (M1-H/M1-H) with untagged M1 dimers (M1/M1). Dissociation and exchange producing M1-H/M1 heterodimers was evident at 37°C but not at lower temperatures (Fig. 3C). Similarly, dissociation of M1-H/M1 heterodimers occurred at 37°C but not at lower temperatures (Fig. 3C). These results indicated that structural instabilities in M1, although dampened at low temperatures, are prominent at physiological temperature.

To investigate the role of structural instability in M1, we focused on the B repeats, owing to their sufficiency for fibrinogen binding. Thirteen substitutions were introduced to set *a* and *d* positions in the B repeats to Val and Leu, respectively, yielding M1\* (residues 42 to 453) and M1<sup>AB\*</sup> (residues 42 to 194) (Fig. 1B and fig. S5A). These substitutions made the core residues optimal for the formation of dimeric parallel coiled coils (12, 24). In addition, we deleted Leu<sup>133</sup> from M1\* and M1<sup>AB\*</sup> [yielding M1\*( $\Delta$ L133) and M1<sup>AB\*</sup>( $\Delta$ L133), respectively] to remove the frameshift in the B repeats (fig. S5B).

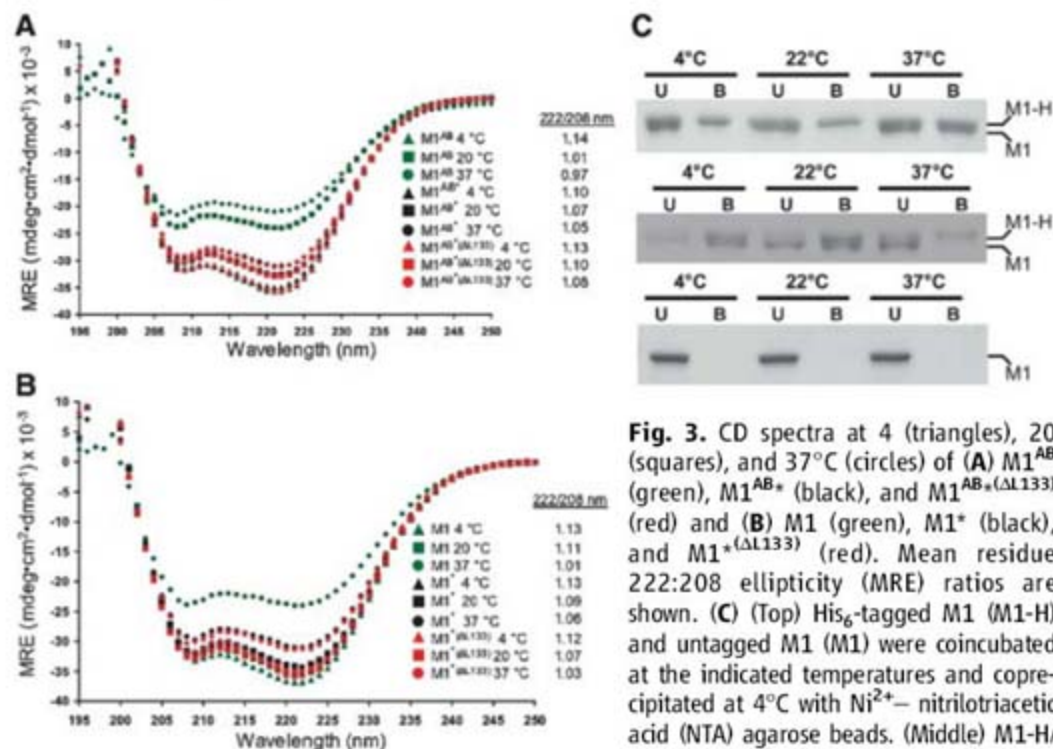
All mutant proteins contained greater  $\alpha$ -helical content as compared with wild-type (WT) proteins at 37°C (Fig. 3). Although enhanced in stability, both M1\* and M1\*( $\Delta$ L133) bound significantly less FgD than did WT M1 at 37°C (Fig. 4, A and B). Binding to human immunoglobulin Gs, an interaction dependent on M1 regions outside the B repeats, was unaffected (fig. S6). Consistent with these results, human neutrophils stimulated with M1\*( $\Delta$ L133) released substantially less HBP as compared with M1

(Fig. 4C). Furthermore, when M1 was injected intravenously into mice, intra-alveolar edema was evident by 30 min in lung histopathologies of 4 out of 4 animals (Fig. 4D), but vascular leakage was absent in all mice injected with M1\*( $\Delta$ L133). M1\*( $\Delta$ L133) did retain some proinflammatory activity, as vascular congestion was comparable for M1 and M1\*( $\Delta$ L133).

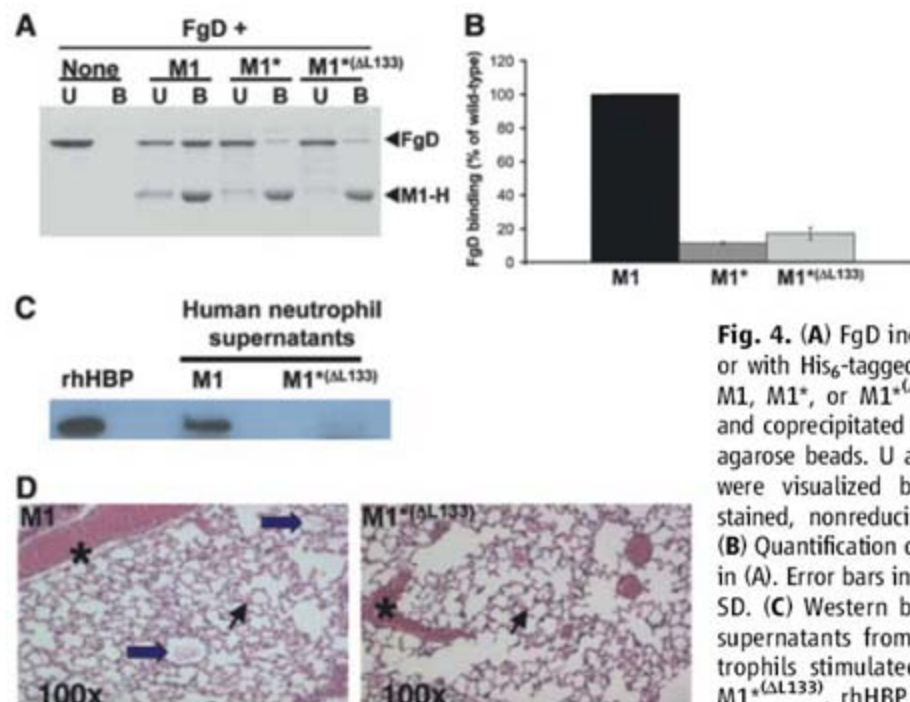
We next examined the cross-reactivity of idealized M1 using an extensively characterized

group of cross-reactive monoclonal antibodies (mAbs) (25). In this group, mAb 36.2.2, which recognizes myosin and tropomyosin and is also highly cytotoxic against heart cells (26), bound M1 most strongly but was 8 to 16 times less reactive against M1\* and M1\*( $\Delta$ L133) (Fig. 5A, fig. S7, and table S2). Thus, sequence idealization of M1 could reduce cross-reactivity.

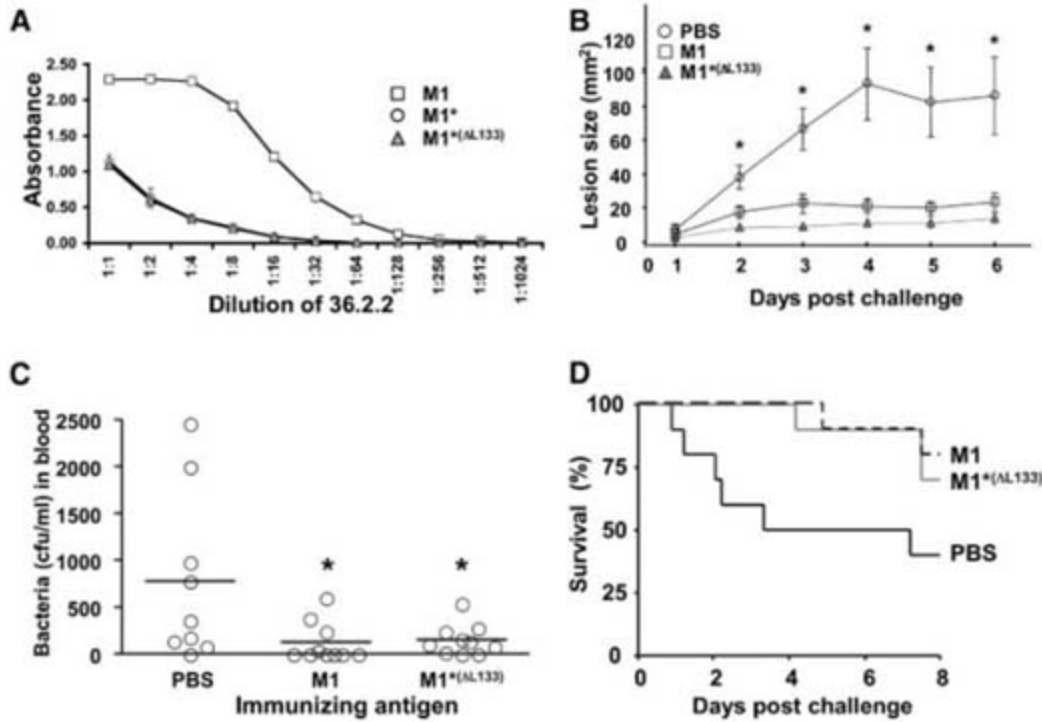
Mice were then immunized with M1 or M1\*( $\Delta$ L133) and challenged with a WT strain of



**Fig. 3.** CD spectra at 4 (triangles), 20 (squares), and 37°C (circles) of (A) M1<sup>AB</sup> (green), M1<sup>AB\*</sup> (black), and M1<sup>AB\*</sup>( $\Delta$ L133) (red) and (B) M1 (green), M1\* (black), and M1\*( $\Delta$ L133) (red). Mean residue 222:208 ellipticity (MRE) ratios are shown. (C) (Top) His<sub>6</sub>-tagged M1 (M1-H) and untagged M1 (M1) were coincubated at the indicated temperatures and coprecipitated at 4°C with Ni<sup>2+</sup>-nitrilotriacetic acid (NTA) agarose beads. (Middle) M1-H/M1 heterodimers were isolated, incubated, and coprecipitated at the indicated temperatures with Ni<sup>2+</sup>-NTA agarose beads. (Bottom) Only untagged M1 was incubated with beads. (A to C) Unbound protein (U) and protein bound to the beads (B) were visualized by Coomassie-stained, reducing SDS-polyacrylamide gel electrophoresis (PAGE).



**Fig. 4.** (A) FgD incubated alone or with His<sub>6</sub>-tagged constructs of M1, M1\*, or M1\*( $\Delta$ L133) at 37°C and coprecipitated with Ni<sup>2+</sup>-NTA agarose beads. U and B proteins were visualized by Coomassie-stained, nonreducing SDS-PAGE. (B) Quantification of FgD binding in (A). Error bars indicate mean  $\pm$  SD. (C) Western blot of HBP in supernatants from human neutrophils stimulated with M1 or M1\*( $\Delta$ L133). rhHBP, recombinant human HBP. (D) Lung histopathology of Balb/c mice 30 min after intravenous injection of M1 or M1\*( $\Delta$ L133). Representative histopathology (hematoxylin and eosin stain) with intra-alveolar edema (thick blue arrows) and macrovascular (asterisks) and microvascular (thin arrows) congestion is indicated. Magnification,  $\times 100$ .



**Fig. 5. (A)** Titer of mAb 36.2.2 versus M1, M1\*, and M1\*(ΔL133) by enzyme-linked immunosorbent assay. Error bars indicate mean ± SD. **(B)** Skin lesion size of mice immunized with M1 or M1\*(ΔL133) after subcutaneous challenge with WT M1 GAS. Error bars indicate mean ± SEM (N = 10 mice per group). Analysis of variance (ANOVA) was significant (P < 0.002) on days 2 to 6; posthoc group comparisons (Tukey-Kramer multiple-comparison test) revealed significant protection of M1 or M1\*(ΔL133) versus phosphate-buffered saline (PBS) on days 2 to 6 (asterisks denote P < 0.05). **(C)** Bacteremia of mice immunized with M1 or M1\*(ΔL133) 4 hours after intraperitoneal challenge with WT M1 GAS. Mean (horizontal bars) and distribution are shown (N = 10 per group). ANOVA was significant at P = 0.02; posthoc group comparisons revealed significant protection of M1 or M1\*(ΔL133) versus PBS control (asterisks denote P < 0.05). **(D)** Kaplan-Meier survival curve of immunized mice from (C).

M1 GAS. M1 and M1\*(ΔL133) elicited similar titers of M1-reactive antibodies (fig. S8), and each afforded similar levels of protection against the development of skin lesions after subcutaneous GAS challenge (Fig. 5B). Similarly, M1 and M1\*(ΔL133) provided comparable levels of protection against acute bacteremia and mortality after intraperitoneal GAS challenge (Fig. 5, C and D).

Our results show that the specific structure of M1 causes proinflammatory interactions with fibrinogen. A comparable set of structural features occurs in myosin and tropomyosin (17–22), indicating a deep level of molecular similarity between M1 and these host proteins and explain-

ing the patterns of cross-reactivity seen in rheumatic fever. Mutation to stabilize the structure of the M1 coiled coil reduced fibrinogen binding, proinflammatory effects, and recognition by a cross-reactive and cytotoxic antibody, whereas it left the immunogenic and protective properties of M1 undiminished.

**References and Notes**

1. M. W. Cunningham, *Clin. Microbiol. Rev.* **13**, 470 (2000).
2. V. A. Fischetti, *Clin. Microbiol. Rev.* **2**, 285 (1989).
3. G. N. Phillips Jr., P. F. Flicker, C. Cohen, B. N. Manjula, V. A. Fischetti, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4689 (1981).
4. U. Ringdahl et al., *Mol. Microbiol.* **37**, 1318 (2000).

5. C. Sandin, F. Carlsson, G. Lindahl, *Mol. Microbiol.* **59**, 20 (2006).
6. S. Chatellier et al., *Infect. Immun.* **68**, 3523 (2000).
7. H. Herwald et al., *Cell* **116**, 367 (2004).
8. B. F. Massell, L. H. Honikman, J. Amezcua, *JAMA* **207**, 1115 (1969).
9. I. Andre et al., *Biochemistry* **45**, 4559 (2006).
10. B. H. Nilson et al., *Biochemistry* **34**, 13688 (1995).
11. B. Tripet, K. Wagschal, P. Lavigne, C. T. Mant, R. S. Hodges, *J. Mol. Biol.* **300**, 377 (2000).
12. P. B. Harbury, P. S. Kim, T. Alber, *Nature* **371**, 80 (1994).
13. M. R. Hicks, J. Walshaw, D. N. Woolfson, *J. Struct. Biol.* **137**, 73 (2002).
14. Materials and methods are available as supporting material on Science Online.
15. S. A. McNeil et al., *Clin. Infect. Dis.* **41**, 1114 (2005).
16. G. Spraggon, S. J. Everse, R. F. Doolittle, *Nature* **389**, 455 (1997).
17. J. H. Brown et al., *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8496 (2001).
18. J. H. Brown et al., *Proc. Natl. Acad. Sci. U.S.A.* **102**, 18878 (2005).
19. W. Blankenfeldt, N. H. Thoma, J. S. Wray, M. Gautel, I. Schlichting, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17713 (2006).
20. Y. Li et al., *Nature* **424**, 341 (2003).
21. Y. Li et al., *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7378 (2002).
22. N. J. Greenfield et al., *J. Mol. Biol.* **364**, 80 (2006).
23. S. Y. Lau, A. K. Taneja, R. S. Hodges, *J. Biol. Chem.* **259**, 13253 (1984).
24. S. C. Kwok, R. S. Hodges, *J. Biol. Chem.* **279**, 21576 (2004).
25. N. M. Mertens, J. E. Galvin, E. E. Adderson, M. W. Cunningham, *Mol. Immunol.* **37**, 901 (2000).
26. M. W. Cunningham et al., *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1320 (1992).
27. This work was supported by NIH grant T32 GM008326 (C.M.), Swiss National Science Foundation fellowship PBZHB-108365 (A.S.Z.), NIH grant R01 AI048694 (V.N.), and American Heart Association and NIH grant R21 AI071167 (P.G.). We thank R. Doolittle, L. Pandi, S. Choe, S. Pegan, G. Allendorph, W. Blankenfeldt, S. Strelkov, G. Ghosh lab members, the staffs of Advanced Light Source 5.0.2 and Advanced Photon Source ID-19, N. Varki, A. Mascaro-Blanco, and S. Mel for reagents, expert technical assistance, or both. Coordinates and structure factors have been deposited in the Protein Data Bank (www.rcsb.org) with the accession number 2OT0.

**Supporting Online Material**

www.sciencemag.org/content/full/319/5868/1405/DC1  
 Materials and Methods  
 SOM Text  
 Figs. S1 to S7  
 Tables S1 and S2  
 References

21 December 2007; accepted 5 February 2008  
 10.1126/science.1154470

# PROTEIN BIOMARKERS SEEK VALIDATION

As proteomics researchers uncover potential disease markers by the hundreds, basic scientists and equipment manufacturers are still struggling with the problem of testing and validating this new trove of results. **By Alan Dove**

In 1896, administrators at Johns Hopkins Hospital in Baltimore approved an extravagance: they bought a suite of specialized equipment for culturing microorganisms, examining tissues, and analyzing body fluids. Stuffing this gear into a 12-foot-square room, the hospital created the world's first dedicated clinical laboratory. The total bill was \$50.

Fed by a rush of new discoveries in microbiology and pathology, clinical laboratories and their specialized tests soon became ubiquitous. Though the field required increasing quantities of expensive equipment and trained technicians, the enormous clinical benefits more than offset the mushrooming costs. Knowing exactly what was wrong with a patient—and knowing it as early as possible—became the benchmark of good medical care.

Like most revolutions, clinical testing soon reached a plateau. Even today, many deadly diseases remain extremely hard to diagnose. If you have strep throat or a thyroid condition, your doctor will know it very quickly, but cancer or Alzheimer's disease might not be apparent until the disease is quite advanced.

Armed with new molecular tools, developed as the field of proteomics has grown, some researchers are now trying to tackle these tough cases with panels of protein biomarkers. Even in the early stages of a disease, the theory goes, cells in the body often change the expression levels of numerous proteins. Detecting those changes would allow clinicians to identify the disease quickly and more definitively. Reducing that theory to practice hasn't been easy, but recent work suggests that it is at least possible.

## Less Is More

The new strategy differs markedly from the old way of developing clinical tests. "The historical trend was you were more focused on a mechanism, and you developed tests for that," says Martin Latterich, associate professor of pharmacy at the **University of Montreal** in Montreal, Quebec. With the advent of high throughput technologies, some researchers started comparing whole proteomes instead, looking at all of the proteins expressed in a sample from a sick patient, and comparing that panel to the proteome of a healthy control. "Instead of looking at one thing at a time, we look at complex samples and don't necessarily worry about if the protein itself is mechanistically involved," says Latterich.

Casting such a broad net, the scientists inevitably find changes that are directly related to the disease mechanism, but they also find many surrogate biomarkers, which change in response to downstream events as the disease progresses. Critics of high throughput biomarker screening argue that this shotgun approach lacks the intellectual rigor of traditional mechanistic studies. Latterich, who responded to that complaint in a 2005 editorial in *Proteome Science* (*Proteome Sci.* 3:8, October 28, 2005), prefers to see mechanistic and high throughput studies as complementary, with different biologists choosing to learn "less about more, or more about less." [continued >](#)



“Instead of looking at one thing at a time, we look at complex samples and don't necessarily worry about if the protein itself is mechanistically involved.”

## Look for these Upcoming Articles

Biomarker Discovery — March 28

Genomics 1 — April 4

RNAi — June 6

*Inclusion of companies in this article does not indicate endorsement by either AAAS or Science, nor is it meant to imply that their products or services are superior to those of other companies.*

## Proteomics

“Validation really requires a joint effort between one lab that would be devoted to discovery and then another lab that has expertise in antibodies or ELISA.”



A more practical problem with modern biomarker screens is that they also uncover numerous false leads. “We’re really going into a lot of these studies blind; we don’t know exactly what the biomarker’s going to be. It’s not a case where we can look at all the differences and say ‘Aha, that’s the biomarker,’” says Timothy Veenstra, director of the laboratory of proteomics and analytical technologies, SAIC-Frederick at the National Cancer Institute (NCI) at Frederick, Maryland.

Indeed, many biomarker researchers cite validation—the process of winnowing a huge panel of potential biomarkers down to a much smaller set of usable ones—as the field’s biggest challenge. For Veenstra, whose lab focuses on biomarkers for prostate, breast, and ovarian cancer, the first problem is deciding which of the hundreds of protein expression changes in an initial screen are even worth pursuing. “Some you can throw out, maybe because they’re related to inflammation or acute phase response proteins that aren’t going to have any specificity for a particular cancer,” says Veenstra.

Even after throwing out obvious red herrings, researchers must still develop robust, reproducible tests for them. “Validation really requires a joint effort between one lab that would be devoted to discovery and then another lab that has expertise in antibodies or ELISA,” says Veenstra. ELISAs, or enzyme-linked immunosorbent assays, are the standard protein-screening tool in clinical labs, but developing a new ELISA test can take months. The more candidate biomarkers a team wants to take into validation, the more tests they need to develop.

To help address this bottleneck, manufacturers are starting to streamline the decades-old ELISA technique. One promising development is Perkin Elmer’s AlphaLISA system. Based on a proprietary antibody-binding bead system, the AlphaLISA eliminates the finicky washing steps of the standard ELISA procedure, and also makes the assay much easier to automate. The latter feature could be a particularly strong selling point as tests based on multiple biomarkers start to filter into already-overtaxed clinical laboratories.

Indeed, the first proteomics-based clinical assays may be just around the corner. Pharmaceutical and clinical testing companies have already shown they are willing to invest in the new generation of biomarkers, even though validation and testing can be costly. After NCI researchers discovered protein expression patterns that appear to correlate with ovarian cancer prognosis, the clinical laboratory firm Correligic Systems of Rockville, Maryland, was quick to pick up the project. The resulting test, known as OvaCheck, remains in regulatory

limbo, but experts seem confident that tests like it will soon reach the market. Meanwhile, the US Food and Drug Administration has started soliciting proteomic data from companies, in order to keep its staff up to speed on the technology as the new proteomics-based tests continue to move into the regulatory process.

### Chip, Column, or Gel?

With numerous pharmaceutical and diagnostic companies working to validate biomarkers with traditional techniques such as ELISA, at least one company thinks it has found a better solution. At Protagen, in Dortmund, Germany, researchers are using protein chips to screen for autoantibody-based biomarkers. In a typical experiment, the scientists apply serum from a patient to a chip with a huge panel of recombinant human proteins, representing the proteome of the fetal brain. “We are comprehensively screening patient sera against this collection of 10,000 proteins, [and] are finding unique autoantibody recognition patterns that can be used as diagnostics,” says Christoph Huels, the company’s CEO.

One example of success using the protein biochip is the ability “to diagnose multiple sclerosis (MS) with more than 80 percent sensitivity and specificity,” according to Huels.

Using a chip-based system for both the discovery and validation phases could help get tests into the clinic faster, but the strategy has limitations. For example, it will reveal only biomarkers that involve autoantibody responses, and only when they target antigens that appear in the company’s panel of recombinant proteins.

Nonetheless, Huels argues that the approach may be more broadly useful than expected. “In the beginning we said it might be only applicable to autoimmune disease, [but] so far in each disease we have looked at it’s working,” he says. Besides MS, the company has found distinctive autoantibody patterns in Alzheimer’s disease and prostate cancer, and academic researchers have proven the principle in alopecia and dilated cardiomyopathy as well.

Though Protagen currently focuses on its own drug development programs, it is not the only one working with autoantibody-based screening chips. Scientists who want to buy the technology off the shelf can simply call Invitrogen, which offers several types of antibody discovery chips in its ProtoArray line. The chips can profile as many as 8,000 proteins in a single run, using as little as 1 $\mu$ L of serum or plasma as a starting sample.

For many gear makers and basic researchers, though, biomarker discovery focuses on more traditional technologies, especially liquid chromatography and mass spectrometry. Fortunately, these tools have been evolving rapidly in recent years (see “Mass Spectrometry for the Masses,” *Science* 319:1115, 2008), and companies have also introduced new equipment designed specifically for biomarker discovery.

One of the first problems biomarker screeners encounter is the chore of depleting the most common proteins from serum or other biological fluids, to allow them to detect the much scarcer but more informative protein signals that might make good biomarkers. While many groups turn to immunodepletion columns for this, Bio-Rad in Hercules, California, recently introduced an alternative, called ProteoMiner.

Instead of antibodies on the column, Bio-Rad uses “a hexapeptide library with diversity, so it’s a combinatorial chemistry–synthesized library,” says Aran Paulus, the company’s research and development manager for proteomics. Most serum proteins will bind to at least one hexapeptide in the column, but once that hexapeptide is saturated, the surplus protein washes through. After washing the column, the researcher can elute the bound proteins, which come out in a much smaller range of concentrations. According to the company, common proteins are depleted across the board, while rare proteins are retained without bias.

Whatever strategy a researcher uses for initial sample preparation, it often requires some additional cleanup before entering the mass spectrometer. For many modern biomarker screeners, that means a nanoscale high performance liquid chromatography system, which is often connected directly to the mass spectrometer. Unsurprisingly, equipment manufacturers are also trying to squeeze more performance out of this phase of the process.

One new strategy is to switch from the usual nanoscale system, which splits the fluid flow in order to achieve the correct flow rate, to a microfluidic pump that regulates its flow rate by feedback. “Our technology allows you to generate these flow rates in the ranges of 50–300 nanoliters per minute without flow splitting, and it provides good reproducibility,” says Remco van Soest, product manager for **Eksigent Technologies** in Dublin, California, which makes a feedback-regulated pump.

The mass spectrometer is the common endpoint for proteomic experiments, but some biomarker studies still perform their initial separations with a much older technique: 2-dimensional gel electrophoresis. Indeed, for many biochemists, 2-D gel technology is akin to the fictitious rock band Spinal Tap—though neither a critics’ nor a public favorite, it continues to fill a much-needed void.

“A lot of people find [the 2-D gel technique] very cumbersome to work with, because it’s long, it’s labor intensive, and it’s believed to be not very reproducible,” says Bio-Rad’s Paulus. “However, in our experience if you have the appropriate training and use the methods that are available, and use them religiously, you get very reproducible results,” he adds.

For those willing to invest the effort, 2-D gels offer distinct advantages over other chromatography techniques in proteomics. The gels preserve posttranslational modifications, which can be especially important in biomarker studies, as many diseases can perturb glycosylation or other modifications. Plus 2-D gels offer tremendous resolving power compared to conventional liquid chromatography, with a single gel run capable of separating thousands of proteins into distinct spots. “No other chromatography technology is coming even close. The resolution of 2-D gels is still unparalleled,” says Paulus.

The technique is also getting at least a little bit easier, thanks to equipment manufacturers who have continued to tweak it. Bio-Rad, for example, offers an extensive line of kits, reagents, and equipment to streamline every step of the 2-D gel process, from sample preparation to gel imaging and protein digestion. Once the proteins are removed from the gel and digested into peptides, biomarker researchers can feed them into a mass spectrometer just like any other sample.

### Hired Help

Considering the complexity of biomarker studies, it’s no surprise that many pharmaceutical companies—and even a few basic researchers—hire someone else to do at least some of the work. Indeed, as the field of biomarker discovery becomes more popular, more companies seem to be wading into the contract business, offering everything from occasional assistance with 2-D gels to complete biomarker discovery and validation services.

“We can carry studies all the way through from start to finish,” says Mike Pisano, CEO of **NextGen Sciences** of Cambridgeshire, UK. Like a handful of other firms, the company has been drawn into the biomarker business partly by the shifting needs of its pharmaceutical company clients.

NextGen started out in laboratory automation, but the rise of biomarker screening led it to acquire Pisano’s original company, Proteomic Research Services of Ann Arbor, Michigan, which specialized in protein purification. Now biomarker work has become a major part of the group’s business. “We developed assays for monitoring proteins quantitatively in clinical samples, [and] what we’ve found is that the pharmaceutical companies are very interested in looking at a set of proteins in clinical trials, in preclinical animal models, or in paraffin-embedded tissues,” says Pisano.

The products that emerge from these experiments may bear little resemblance to traditional clinical tests. Because they will rely on panels of proteins rather than simple assays, for example, interpretation could be more complicated. “From panels, in order to come to some meaningful conclusion, you have to classify the type of patient groups that have a certain type of pattern, and to link that to disease outcome or treatment option,” says Montreal’s Latterich. Equipping the next generation of clinical labs to do those types of analyses, of course, will probably cost more than \$50.

*Alan Dove is a science writer and editor based in Massachusetts.*

DOI: 10.1126/science.opms.p0800023

### Featured Participants

**Bio-Rad**  
[www.bio-rad.com](http://www.bio-rad.com)

**Perkin Elmer**  
[las.perkinelmer.com](http://las.perkinelmer.com)

**Correlogic Systems**  
[www.correlogic.com](http://www.correlogic.com)

**Protagen**  
[www.protagen.de](http://www.protagen.de)

**Eksigent Technologies**  
[www.eksigent.com](http://www.eksigent.com)

**SAIC-Frederick**  
[www.saic.com/life-sciences](http://www.saic.com/life-sciences)

**Invitrogen**  
[www.invitrogen.com](http://www.invitrogen.com)

**University of Montreal**  
[www.umontreal.ca/english](http://www.umontreal.ca/english)

**NextGen Sciences**  
[www.nextgensciences.com](http://www.nextgensciences.com)



## New Products

**Biophotometer**

Building on the strengths of the UV/Vis BioPhotometer, the BioPhotometer Plus features three additional wavelengths, a new keypad design, and the ability to use microliter cells as well as cuvettes. This compact photometer is easy to operate and produces reliable, reproducible results quickly. The three new wavelengths increase the number of routine methods available to 32, compared with 12 for the earlier BioPhotometer; nine of the methods are freely programmable, including measurement of DNA, RNA, and protein concentration; enzyme assays; and the incorporation rate of fluorescent dyes. Additional methods available include the incorporation rate and concentration of fluorescent dyes in nucleic acids or proteins, measurement of enzymatic activities, cell biological assays, and direct measurement at single wavelengths. Both measurement and results calculations, including factoring in dilutions, are performed at the push of a button. Related methods are grouped for easy selection on the redesigned keypad, and differently colored function keys reduce operator errors.



**Eppendorf**

For information +49 40 53801 0  
www.eppendorf.com

**Rapid Conjugation Kits**

The Lynx Rapid Conjugation Kits need just 30 seconds handling time for labeling primary antibodies with a wide range of fluorochromes and enzymes. The one-step labeling technology offers high conjugation efficiency with 100 percent antibody recovery. The kits are suitable for both large and small quantities of antibodies. The Lynx labels are covalently and directionally conjugated to the user's antibody in a controlled process at near-neutral pH, requiring no subsequent desalting steps. This offers significant advantages over traditional labeling methods that need subsequent sample concentration and dialysis, in which antibody loss is a risk and batch-to-batch consistency for experiments cannot be guaranteed. The hands-on time remains 30 seconds whether the user is conjugating microgram or milligram quantities of antibody, so a researcher can do trial runs on conjugated antibodies to optimize the application protocols on a small scale before easily scaling up.

**AbD Serotec**

For information +44 1865 852 700  
www.ab-direct.com

**Complex Peptide Separations**

The NanoLC-Ultra, a new proteomics system that delivers gradient precision at high pressures of up to 10,000 psi, is suitable for the chromatographic separation of complex peptide mixtures. It can be used with longer columns than previous systems and also accommodates columns packed with smaller particles. The NanoLC-Ultra features self-priming and self-purging pumps for easy solvent changes as well as a temperature-controlled column compartment that maintains retention time reproducibility. It can operate in a "peak parking" mode with the flow rate temporarily decreased, to facilitate longer mass spectrometry interrogation of the analyte. By eliminating flow splitting, the proprietary microfluidic flow control technology generates accurate and reproducible flow rates with a resolution of more than 1 nl per minute. The control software includes integration drivers for the most widely used mass spectrometry programs.

**Eksigent**

For information 925-560-2637  
www.eksigent.com

**Protein Separation Unit**

The MF10 Protein Separation Unit simultaneously separates proteins based on size and charge. The technique makes use of a combination of polyacrylamide membranes with a gradation of pore sizes and electrophoretic charges through the buffer. Researchers can choose from a selection of color-coded molecular weight cut-off membranes spanning the range of proteins under investigation. By controlling membranes, buffer pH, and voltage, the MF10 can create up to eight

fractions simultaneously on samples as small as 200 microliters. The two-dimensional polyacrylamide membrane separation technology allows fraction preparation, desalting, and depletion of abundant proteins, upstream of two-dimensional gels or mass spectrometry. The method is nondestructive, so proteins retain their biological activity. The small, affordable system is designed to fill a gap in the proteomics tool arsenal by allowing individual researchers to do a variety of small-scale protein preparations.

**NuSep**

For information 908-575-1378  
www.nusep.com

**Mass Spectrometer**

The LTQ Orbitraps are hybrid linear ion trap mass spectrometers. The LTQ Orbitrap Discovery is a cost-effective system suited for general proteomics and metabolism applications, offering an alternative to time-of-flight systems. Featuring high resolution and accurate mass capabilities, this hybrid liquid chromatography/mass spectrometry system achieves high-throughput, label-free profiling for differential expression while the mass spectrometry capability makes it suitable for small molecule characterization and structural elucidation. The Discovery offers up to 30K resolution power and an exceptional scan rate, enabling the separation of trace-level, isobaric analytes. The LTQ Orbitrap XL is a protein identification and biomarker discovery platform that features sensitivity, rapid scan rate, high mass accuracy, and up to 100K resolution power. The Orbitrap XL features a new collision cell for increased flexibility in fragmentation applications including peptide quantitation, de novo sequencing, and metabolomics research. It is also upgradeable to use the electron transfer dissociation option for controlled dissociation of peptides and proteins.

**Thermo Scientific**

For information 800-532-4752  
www.thermo.com/ms

Electronically submit your new product description or product literature information! Go to [www.sciencemag.org/products/newproducts.dtl](http://www.sciencemag.org/products/newproducts.dtl) for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.